COVARIATE FACTOR MITIGATION TECHNIQUES FOR ROBUST GAIT RECOGNITION

by

Tenika P. Whytock

Submitted for the degree of
Doctor of Philosophy

INSTITUTE OF SENSORS, SIGNALS AND SYSTEMS
SCHOOL OF ENGINEERING AND PHYSICAL SCIENCES
HERIOT-WATT UNIVERSITY

May 2015

Abstract

The human gait is a discriminative feature capable of recognising a person by their unique walking manner. Currently gait recognition is based on videos captured in a controlled environment. These videos contain challenges, termed *covariate factors*, which affect the natural appearance and motion of gait, e.g. carrying a bag, clothing, shoe type and time. However gait recognition has yet to achieve robustness to these covariate factors.

To achieve enhanced robustness capabilities, it is essential to address the existing gait recognition limitations. Specifically, this thesis develops an understanding of how covariate factors behave while a person is in motion and the impact covariate factors have on the natural appearance and motion of gait. Enhanced robustness is achieved by producing a combination of novel gait representations and novel covariate factor detection and removal procedures.

Having addressed the limitations regarding covariate factors, this thesis achieves the goal of robust gait recognition. Using a skeleton representation of the human figure, the Skeleton Variance Image condenses a skeleton sequence into a single compact 2D gait representation to express the natural gait motion. In addition, a covariate factor detection and removal module is used to maximise the mitigation of covariate factor effects. By establishing the average pixel distribution within training (covariate factor free) representations, a comparison against test (covariate factor) representations achieves effective covariate factor detection. The corresponding difference can effectively remove covariate factors which occur at the boundary of, and hidden within, the human figure.

To my mum and dad

Acknowledgements

First to my supervisors Alexander Belyaev and Neil Robertson. Thank you for your continual patience and encouragement during my journey. You have helped evolve my research skills and confidence in pursuing my curiosity.

I would like to thank my examiners Paul Siebert and Kartic Subr for their constructive comments after extensively reading my thesis.

I am particularly thankful for the financial support of the ESPRC, British Machine Vision Association and the Institute of Signals, Sensors and Systems. This has enabled me to pursue my personal goal of obtaining a PhD, as well as travel nationally and internationally to present my research.

Finally to my parents. A special thank you for your constant love, support and cake. All of which have been essential during my undergraduate and postgraduate studies at Heriot-Watt University.

Publications

This thesis has yielded the following publications:

1. T.P. Whytock, A. Belyaev, N.M. Robertson. GEI + HOG for Action Recognition. *British Machine Vision Conference Student Workshop*, 2012

2. T.P. Whytock, A. Belyaev, N.M. Robertson. Improving Robustness and Precision in GEI + HOG Action Recognition. *Advances in Visual Computing, Lecture Notes in Computer Science, Part I*, Volume 8033, pp. 119-128, 2013 (Presented at the International Symposium on Visual Computing)

3. T.P. Whytock, A. Belyaev, N.M. Robertson. Towards Robust Gait Recognition. *Advances in Visual Computing, Lecture Notes in Computer Science, Part II*, Volume 8034, pp. 523-531, 2013 (Presented at the International Symposium on Visual Computing)

4. T.P. Whytock, A. Belyaev, N.M. Robertson. Robust Gait Recognition Via Covariate Factor Mitigation. *International Conference on Imaging for Crime Detection and Prevention*, 2013

5. T.P. Whytock, A. Belyaev, N.M. Robertson. Dynamic Distance-based Shape Features for Gait Recognition. *Journal of Mathematical Imaging and Vision*, Volume 50(3), pp. 314-326, 2014

6. T.P. Whytock, A. Belyaev, N.M. Robertson. On Covariate Factor Detection and Removal for Robust Gait Recognition. *Machine Vision and Applications* (to appear), 2015

Contents

Chapter 1

Introduction

This thesis has developed gait recognition approaches with enhanced robustness capabilities. This is achieved by understanding a) how covariate factors behave while a person is in motion and b) the impact covariate factors have on the natural appearance and motion of gait. As a consequence of increased robustness, gait recognition can progress towards validation in more complex and unconstrained datasets. Ultimately, this is another step towards using gait as a biometric in the real world.

Identifying a person is essential for everyday life such as financial transactions, travel and security. There are multiple formal means of identification, e.g. a birth certificate, passport, driving license and bank cards, all of which tend to be verified by a photograph, signature, password or PIN number. However these verification means are by no means foolproof. Biometrics are an alternative means of identification which are difficult to fake, disguise and forget. Considering bank fraud, Barclays' is to become the first UK bank to deploy the finger vein biometric for business banking customers. This removes the need to authorise payments via PIN number, password or authentication code. However despite the Hollywood fantasy, Barclays' CEO has stressed the fact that a severed finger would not fool the technology as the veins would become invisible to infrared light. Biometrics commonly rely on the cooperation with a person of interest to extract reliable data to determine identity. In addition, some biometrics require intrusive data collection. An unobtrusive biometric which does not require cooperation is gait, i.e. the unique walking manner of a person.

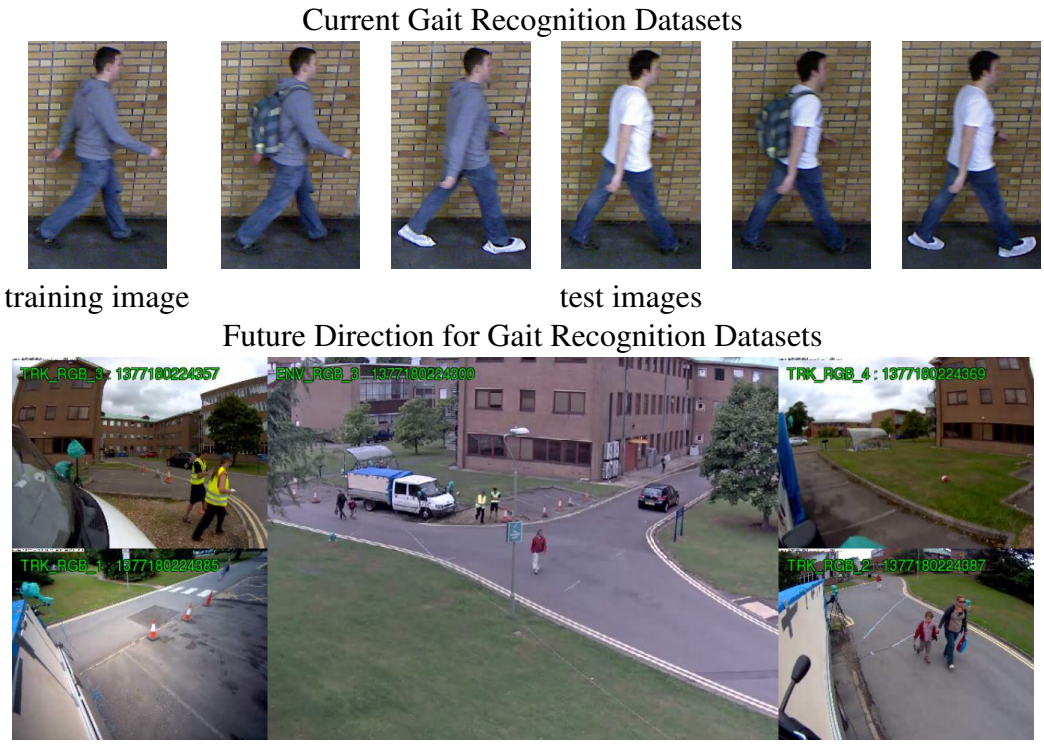


Figure 1.1: The gait recognition problem is currently focussing on somewhat controlled image sequences. Once robustness has reached a satisfactory level, gait recognition will evolve to validation based on unconstrained videos "in the wild"

Gait was shown to be unique in medical research and psychophysics research during the 1960s and 1970s. Therefore gait recognition exploits the unique nature of gait to identify a person. Now gait recognition is a research topic which has attracted the attention of numerous commercial and academic institutions across the world. There are numerous real world applications that can benefit from gait recognition, e.g. visual surveillance, forensics, and robotics to name a select few. As such, gait recognition is a very active and highly competitive research field.

Compared to established biometrics such as fingerprint, gait is a relatively young biometric. As such, validation is based on dataset image sequences seen in Figure 1.1. While these images appear simplistic, it is essential for gait recognition to establish a degree of robustness during such single person per image sequence datasets. Despite the simplistic nature, these datasets use numerous real world challenges which can alter the natural appearance and motion of gait. The challenges, termed covariate factors by the gait recognition research community, include clothing and carrying a bag which can be seen in Figure 1.1. Research identifies the ultimate goal for gait recognition validation which is based on complex, unconstrained real world image sequences such as those seen in Figure 1.1.

Note that the majority of research considers gait recognition as a non-real time problem. However, the ultimate goal for gait recognition is real-time processing.

1.1 Gait Recognition Development

Humans have demonstrated a natural curiosity about motion. The 15th Century sketch books of da Vinci state

> "it is indispensable for a painter, to become totally familiar with the anatomy of nerves, bones, muscles, and sinews, such that he understands for their various motions and stresses, which sinews or which muscle causes a particular motion."

Such anatomical studies progressed to topics including biomechanics (17th Century), cinematography (19th Century) and motion perception (20th Century) before emerging into the computer vision techniques widely employed today.

Finally, it is interesting to read about the discriminative nature of gait reflected in fictional literature such as Shakespeare's *Henry IV/II*

> "For that John Mortimer . . . in face, in gait in speech he doth resemble"

The Tempest

> "High′st Queen of state, Great Juno comes; I know her by her gait"

and *Twelfth Night*

> "By the colour of his beard, the shape of his leg, the manner of his gait, . . . , he shall find himself most pleasingly personated."

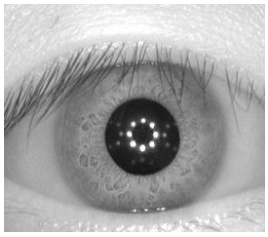
1.2 Biometrics

Biometrics can identify a person based on their characteristics or traits. It is imperative that biometrics are *i*) present in every person, *ii*) capable of differentiating between persons, *iii*) time-invariant to some degree and *iv*) measured quantitatively. For a biometric to

Year	Type
1960	fingerprint, voice
1970	palmprint, face
1980	iris, signature
1990	vein, gait, ear, keystroke
2000	DNA, EEG, dental, shoe


gait


ear


iris


fingerprint

Table 1.1: Biometric modality development, adapted from Nixon et al. (2010)

be deployed, it must be *i*) accurate, *ii*) willingly used by the public and *iii*) robust. While the ideal biometric does not exist, compromises can be taken or biometric fusion can be achieved to yield favourable results. Table 1.1 shows a timeline of biometric modality development; see Yam and Nixon (2009a) for additional information regarding biometric development.

Two distinct biometric modality classes exist, namely physiological and behavioural; see Jain et al. (2004) for additional information. Physiological biometrics are derived from the body and include ear, face, fingerprint, iris, palmprint, retina and DNA. Behavioural biometrics are derived from behavioural patterns including gait, signature, voice and keystroke. Biometrics are limited by the cooperation required for data collection and the distance to the person of interest. Considering identifying a person from a video, it is impossible to interact with a person of interest which renders a number of biometrics ineffective. An excellent candidate for identifying a person from a video is gait. Gait is an effective biometric given *i*) no cooperation or consent is required and *ii*) unobtrusive capture at a distance and low resolution.

Forensics is a major application for biometrics where fingerprint and DNA are traditionally used as they are commonly left at a crime scene. Another vital application is surveillance. Crime rates in the UK have prompted a rapid deployment of close-circuit television (CCTV) surveillance for crime detection and prevention in a bid to provide a safer environment. CCTV, if correctly positioned, can capture a criminal entering or

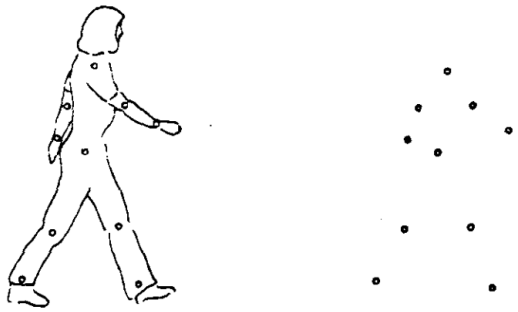


Figure 1.2: Human motion analysis using markers attached to specific locations on the body [Johansson (1973)]

fleeing a crime scene which provides first hand evidence for prosecution. Note that the validity of raw CCTV footage in a court environment is dependent on CCTV video quality. In this situation, facial biometrics are a plausible option for person identification provided that the distance from camera to person be small enough to yield an image proving identity beyond reasonable doubt. However when facial recognition is not plausible, gait as a biometric comes into its element.

Gait as a Biometric

The gait of a person is achieved through a joint effort of the skeleton and muscles working in harmony (for a healthy gait). While the walking pattern is similar across healthy persons, subtle variations attributed to magnitude and timing, i.e. walking manner and posture [Matovski et al. (2013)], yield a unique gait. Early medical research [Murray et al. (1964)] and psychophysics research [Cutting and Kozlowski (1977)] has demonstrated that gait is unique. These early gait analysis approaches rely on markers or lights attached to specific locations on the body [Johansson (1973)] seen in Figure 1.2. When the room is darkened, the human figure is represented solely by a configuration of lights. When the person stands still, it is easy to confuse the lights with a constellation. However when in motion, Cutting and Kozlowski (1977) and Kozlowski and Cutting (1977) shows that humans can easily determine the identity (58% accuracy) and gender (70% accuracy) of a friend. Marker-based approaches are currently used for applications such as clinical gait analysis and motion capture. The marker-based approaches are infeasible, intrusive and impractical for gait recognition applications such as surveillance. Therefore this thesis considers marker-free approaches for gait recognition.

Replicating the human ability to recognise a person by their gait, and more importantly extending the capability to recognising hundreds (minimum) of different people, is a complex computer vision task. Gait is an effective biometric as it is difficult to fake, disguise and forget. In addition, gait can be used to identify a person when traditional biometrics fail, e.g. facial recognition is ineffective when a person is captured at a distance. The effect on performance when a person imitates the gait of another person is unknown, and it is important to investigate this topic to understand the potential vulnerabilities of gait recognition; this is outwith the scope of this thesis.

Compared to the well known and established biometrics, such as fingerprint and face, Table 1.1 shows that gait as a biometric is relatively young. Nevertheless, gait recognition has gained a significant research community in the last decade due to the numerous real world and vital applications. Niyogi and Adelson (1994) are commonly cited as the first to achieve gait recognition via computer vision. The years following have seen approaches develop to match the complexity of current validation datasets. Gait recognition can be used in an online or offline manner where applications include visual surveillance, video indexing, access control, forensics, human interaction, robotics and monitoring the elderly or children. However there are challenges, termed covariate factors by the gait recognition community, which can alter the natural appearance and motion of gait. Examples of covariate factors include clothing, carrying a bag and shoe type.

Finally, be aware that human gait can reveal cues which are discriminative for applications such as age recognition [Lu and Tan (2010),Makihara et al. (2011)] and gender recognition [Li et al. (2008),Yu et al. (2009)], however these topics are outwith the scope of this thesis. Troje (2002) presents a fascinating visual demonstration of the differences between males and females walking and running, and the effect of weight and mood on walking.

Forensic Gait Analysis

Gait has contributed evidence to convictions nationally and internationally [Nixon et al. (2010)], including *i*) the murder case of Swedish Foreign Minister Anna Lindh, *ii*) a bank robber in Noerager, Denmark and *iii*) a burglar in Lancashire, United Kingdom

[Bouchrika et al. (2011)]. These cases promote the use of gait as a biometric to provide evidence for prosecution. However this evidence is currently based on a human expert witness performing the identification via the relatively recent field of forensic podiatry [DiMaggio and Vernon (2011)] and biomechanics. This motivates the research community to provide an automated computer vision alternative i.e. gait recognition. In turn, this could use image-based evidence which has been previously rejected due to well established biometrics, such as face, failing to provide identification. Forensic gait analysis has the potential to implicate, as well as eliminating suspects from enquiries. This process can aid suspects admitting guilt and thus save the cost of going to trial. Most importantly, forensic gait analysis sends a strong message to criminals that concealing their face will not protect them from criminal conviction.

Recognising a person by their gait is faster and more economical compared to traditional methods such as fingerprint and DNA analysis; this is important given the current financial climate and should motivate gait recognition researchers. CCTV has been rapidly deployed in the UK, however the UK public are somewhat unconvinced about the success for crime detection and prevention. In defence, there is only a small proportion of CCTV data which shows interesting or anomalous persons and activities compared to the copious amounts of data collected daily. An interesting tactic employed by the Metropolitan Police Service involves human "super-recogniser" officers who have the ability to recall the faces of hundreds of offenders. This was an effective tool which significantly increased the identification of suspects in the London Riots in August 2011. Unfortunately such officers are not regularly responsible for the detection of crimes or persons of interest in CCTV daily. However, using computer vision in harmony and simultaneously can provide the capability of siphoning the interesting images for more in depth analysis by such officers and/or computers. This "human in the loop" approach combines the unique advantages of humans and computers, but can also yield a streamlined process.

Equine Motion Analysis

There is a considerable amount of research focussed on human motion analysis, however an interesting and lucrative research direction is based on equine motion analysis. Marey

and Muybridge promoted equine locomotion research during the 1920s and 1930s. Interestingly, Muybridge helped answer the highly debated question of whether a trotting or galloping horse ever lifts all four feet completely off the ground - Muybridge demonstrated this to be true. This question was raised by the ancient Egyptians but remained unanswered as the human eye is too slow to decompose the motion of a horse at high speeds. High-speed cameras were commonly used for motion analysis, however more recently gyro-sensors, 3D accelerometers and Global Positioning System data capture are being used. This equipment has replaced kinematics based on cameras and markers. As a consequence, equine motion analysis has moved from the confines of a laboratory and treadmill environment to a more realistic outdoor environment. Equine motion analysis is effective for applications such as diagnosing equine lameness (abnormal gait pattern) and monitoring the recovery after an injury. An interesting extension could determine if the gait of a horse, or indeed any animal, is unique like human gait.

1.3 Aim of this Thesis

This thesis is motivated by the requirement to develop gait recognition approaches with enhanced robustness capabilities to covariate factors. This research will help push gait recognition towards validation in more complex and unconstrained datasets, and therefore take an important step towards ultimately deploying gait as a viable biometric for the numerous real world applications research considers.

The majority of validation datasets present a very select, but common and real world set of covariate factors, namely clothing, carrying a bag, elapsed time between capture, viewpoint and complex couples thereof. While the gait recognition community is tackling the issue of robustness to covariate factors, the limitations of existing gait recognition approaches are *i*) underestimating how covariate factors behave when a person is in motion and *ii*) neglecting the unique impact covariate factors have on the natural appearance and motion of gait. There are numerous ongoing debates for gait recognition implementation. This thesis explores *i*) model-based, model-free and multi-information fusion approaches, *ii*) the number of images in the gait cycle used to represent gait, *iii*) gait representations

and *iv*) techniques to improve robustness. The limitations of existing state-of-the-art gait recognition approaches are established and form the motivation of this thesis.

This thesis considers non-real time gait recognition for a surveillance-type application. The validation datasets contain image sequences which are somewhat simplified by the following assumptions

- all image sequences capture a person walking from a side view due to the visibility of discriminative limb-based motion
- all image sequences contain one person walking only to avoid confusion during periods of occlusion
- all image sequences capture full body views as gait is a full body movement

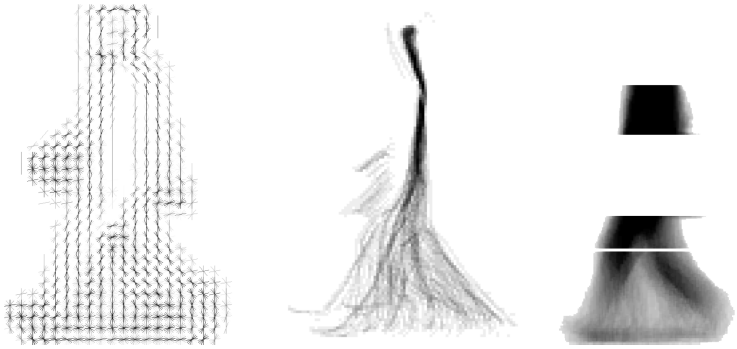
While these assumptions simplify the gait recognition problem, this is necessary as gait recognition must establish tangible robustness solutions at this complexity level prior to developing an equal performance during "unconstrained" image sequences.

This thesis considers the average performance across covariate factor sequences in each validation dataset as this shows covariate factor generalisation capabilities (this is standard in gait recognition). However the performance achieved during individual covariate factor sequences is also important as this identifies limitations in the deployed gait recognition approach.

The goal of this thesis is to progress towards gait recognition "in the wild", and this requires significant performance improvements during the presence of covariate factors. This will be achieved by *i*) establishing how covariate factors behave when a person is in motion and *ii*) determining how the natural appearance and motion of gait is altered when a covariate factor is present.

1.4 Contributions

This thesis presents a combination of innovative techniques and tangible solutions to the problem of enhancing gait recognition robustness. The contributions are stated and linked to the relevant publications. The corresponding gait representations are seen in Figure 1.3.



HOG describing GEI
Chapter 3

SVIM
Chapter 4

CFDR module
Chapter 5

Figure 1.3: Novel techniques to represent human gait robustly

Gait Energy Image Described by Histograms of Oriented Gradients

Presented at the British Machine Vision Conference Student Workshop [Whytock et al. (2012)] and the International Symposium on Visual Computing [Whytock et al. (2013a)].

Hypothesis

This chapter argues that to ensure robustness, HOG parameters (gradient scheme, cell size and bin size) must undergo re-evaluation when applied to applications other than person detection.

Histograms of Oriented Gradients (HOG) is a highly cited feature descriptor. However HOG is commonly employed as a "black box" meaning the parameters are optimal for person detection. HOG is used to describe the Gait Energy Image (GEI) which is a single compact 2D human gait and action representation. Wilcoxon tests indicate that the cell size and bin size significantly affect the action recognition results, while gradient scheme and cell size significantly affect the gait recognition results. The GEI and HOG combination is effective for action recognition when validated in the Weizmann Action dataset. However the GEI and HOG combination does not yield a satisfactory performance for gait recognition when validated in the CASIA B and TUM GAID dataset as *a*) the combination does not scale with dataset size and *b*) HOG encodes the appearance and motion of covariate factors in the GEI.

Variance-based Fuzzy Skeletal Features

Published in the Journal of Mathematical Imaging and Vision [Whytock et al. (2014)].

Hypothesis

This chapter argues that by exploiting the Poisson equation to construct a smooth distance function, fuzzy skeletons can be extracted and formed into a single compact 2D gait representation to yield a discriminative gait descriptor.

The Skeleton Variance Image (SVIM) explores the gap in knowledge relating to the novel combination of skeleton representations and single compact 2D gait representations. A screened Poisson equation is used to define a smooth distance function which absorbs boundary noise given the tunable smoothing parameter. The fuzzy skeleton extracted from the smooth distance function is effective for covariate factor motion mitigation. Discriminative gait motion features are extracted when the fuzzy skeleton sequence is condensed into the SVIM by computing the pixel-wise variance. The SVIM achieves a 9.9% increase over state-of-the-art in the TUM GAID dataset.

Covariate Factor Detection and Removal

Presented at the International Symposium on Visual Computing [Whytock et al. (2013c)] and the International Conference on Imaging for Crime Detection and Prevention [Whytock et al. (2013b)], and published in the Journal of Machine Vision and Applications [Whytock et al. (2015)].

Hypothesis

This chapter argues that single compact 2D gait representations can achieve superior robustness when performing dedicated covariate factor detection and removal.

The covariate factor detection and removal (CFDR) module detects and removes covariate factors in single compact 2D gait representations. Covariate factors are detected by establishing the pixel intensity distribution in covariate factor free training sequences. A degree of tolerance is included to incorporate the natural inter-class and intra-class variance in human gait. This process is effective for differentiating between natural gait motion and

covariate factor motion. Covariate factor motion is achieved by removing complete rows where covariate factors are detected. This process can remove covariate factors which occur at the boundary of, and hidden within, the figure. By applying the CFDR module to the SVIM, a further 3.6% increase over state-of-the-art is achieved in the TUM GAID dataset.

1.5 Thesis Roadmap

The three aforementioned contributions form three distinct and self-contained chapters, where each explicitly states the current limitation of research being addressed. As such, these chapters contain their own conclusion and future directions for development. Therefore the thesis is organised as follows.

Chapter 2 reviews the existing research on gait recognition, as well as the analogous topic of action recognition given these tasks are closely related in the overarching topic of human motion analysis. Covariate factors encountered in gait recognition and action recognition, as well as the standardised datasets available for validation are discussed. The limitations of existing gait recognition research are explicitly defined in order to point towards the research directions taken within this thesis.

Chapter 3 performs a novel qualitative and quantitative evaluation to demonstrate the limitations of employing Histograms of Oriented Gradients (HOG) parameters tuned for person detection. This evaluation is also necessary given the alternative Gait Energy Image action/gait representation and the multi-class action recognition and gait recognition applications. The evaluation for action recognition and gait recognition is based on *i*) eight gradient schemes with varying gradient orientation and gradient magnitude accuracy and *ii*) 100 cell size and bin size combinations. The optimal parameters for person detection are compared against optimal parameters for action recognition and gait recognition.

Chapter 4 promotes skeleton representations for gait recognition as they are infrequently used due to boundary noise sensitivity. Three Poisson-based smooth distance functions are evaluated to determine the accuracy level and degree of smoothness re-

quired for robust gait recognition. The Skeleton Variance Image (SVIM) is formed from condensing the fuzzy skeleton sequence into a single compact 2D gait representation when computing the pixel-wise variance. The SVIM is compared against analogous single compact 2D gait representations.

Chapter 5 maximises covariate factor detection and removal in single compact 2D gait representations with the covariate factor detection and removal (CFDR) module. Covariate factor detection is achieved by analysing the pixel distribution in covariate factor free sequences. The 3-sigma rule is used to apply a degree of tolerance which incorporates the natural inter-class and intra-class variance in the human gait. These processes are used to minimise the pixel-wise confusion between natural gait motion and covariate factor motion. Three covariate factor removal techniques are evaluated to determine the effectiveness in removing covariate factors which occur at the boundary of, and hidden within, the figure. The performance of gait representations is compared with and without the CFDR module applied.

Chapter 6 summarises the contributions of the thesis and provides some interesting future directions for gait recognition.

Chapter 2

Related Work

This chapter provides an overview of state-of-the-art gait recognition research, alongside an insight into the closely related topic of action recognition; these topics are part of the overarching field of human motion analysis. The challenges, termed covariate factors, and standardised datasets used for validation are discussed. The limitations of state-of-the-art gait recognition research prompt the motivations of this thesis.

2.1 Gait Recognition

Li (2009) defines gait as

> "the manner of a person's movement, specifically during walking."

Therefore gait recognition is defined as recognising a person by their walking manner. Note that gait and motion differ; motion describes the action of moving, however gait describes how the action is performed. Gait recognition is an active and competitive research topic in commercial and academic institutions. Researchers are united in providing tangible solutions to numerous real world applications, such as visual surveillance for crime detection and prevention. Regardless of the application, successful gait recognition requires discriminative feature extraction and robustness to real world covariate factors e.g. clothing and carrying a bag.

Gait recognition is closely related to the classical study of human motion analysis [Gavrila (1999),Wang et al. (2003),Hu et al. (2004)], however modern gait recognition

surveys, such as Wang et al. (2010) and Chai et al. (2011), reflect the current debates in gait recognition. Alternative gait recognition techniques [Gafurov (2007)], not necessarily suited towards applications such as surveillance, include *i*) wearable sensors, *ii*) floor pressure sensors, *iii*) Doppler sonar and *iv*) acoustic gait recognition [Hofmann et al. (2013)].

Since the early computer vision attempts by Niyogi and Adelson (1994) and Cunado et al. (1997), gait recognition has significantly developed and numerous implementation debates exist. This section examines the *1*) overarching debate of model-based, model-free and multi-information fusion approaches, *2*) number of images in the gait cycle used to represent gait, *3*) gait representations and *4*) techniques to improve robustness. Gait recognition approaches follow a typical path: gait representation, feature extraction, dimensionality reduction and classification. Note that this section focusses on gait representations and features. Gait classification techniques are not discussed in this section, however surveys by Wang et al. (2010) and Chai et al. (2011) are recommended for additional reading.

2.1.1 Gait Recognition Approaches

A well debated area, and one of the first decisions to be made when using gait recognition, is the choice between model-based, model-free and multi-information fusion approaches.

Model-based approaches focus on human body structure and tend to track or model body segments such as the head, arms and legs which are extracted via anthropometric data [Drillis and Contini (1966),Dempster and Gaughran (1967)]; see Yam and Nixon (2009b) for further reading on model-based approaches. Model-based approaches can extract static features and dynamic features which are more meaningful for humans. For example, static features include body segment and stride lengths, while dynamic features include joint angle trajectories of the lower limbs. There are a number of models available which may be applied to gait, such as ellipses [Lee and Grimson (2002)], stick figures [Johnson and Bobick (2001)], ribbons [Leung and Yang (1995)], blobs [Wren et al. (1997)], pendulums [Yam et al. (2004)] and 3D [Seely (2010)].

Model-free approaches disregard human body structure and instead consider fea-

tures such as the appearance (pose) and motion of gait. Silhouette representations are common as colour and texture are disregarded thus avoiding bias to appearance which is inconsistent over time. Silhouettes are relatively straightforward to extract via sources including Lidar, Time-of-flight and Microsoft Kinect [Hofmann et al. (2012)]. Alternative model-free approaches are derived from silhouettes, such as contours [Wang et al. (2012)], optical flow [Bashir et al. (2009b)], and 3D [Sivapalan et al. (2011)]. Note that the silhouette quality is often a deciding factor for the chosen approach.

Multi-information fusion approaches attempt to mimic human vision perception through feature-level or decision-level fusion. A fusion approach is effective to increase the performance of unique features or biometrics. Feature-level fusion commonly combines appearance (pose) and motion features [Wang et al. (2004)], while decision-level fusion can combine biometrics e.g. gait and face [Hofmann et al. (2012)]. An interesting and recent example of biometric fusion is the Southampton Multi-Biometric Tunnel [Seely et al. (2008),Nixon et al. (2010)] which fuses gait, face and ear to identify a person.

Discussion. Model-based approaches are effective for view, scale and rotation invariance. However these approaches are limited by the *i*) sensitivity to image quality and noise, where higher resolution images are required for accurate models and *ii*) increased computational demands due to the typically high dimensional parameter space. These factors suggest that model-based approaches are less suitable for applications such as surveillance which capture a person at a distance. Cheaper computing power and equipment such as the Microsoft Kinect, which can extract depth information and a skeleton, have increased the use of model-based approaches for gait recognition. While the immediate focus of this thesis is single feature and biometric gait recognition, subsequently proposed research may achieve a higher performance during fusion with other features or biometrics; however this is outwith the scope of this thesis.

Model-free approaches lack invariance to view, scale and rotation. However these approaches are effective due to the *i*) insensitivity to image quality and noise and *ii*) reduced computational demands. These factors are advantageous given gait recognition is concerned with capturing a person at a distance, therefore this thesis focusses on model-free approaches.

Figure 2.1: The gait cycle is defined as the period taken between consecutive heel strikes from the same leg

Single Compact 2D Gait Representations	Year
Motion Silhouette Image Lam and Lee (2005)	2005
Gait Energy Image Han and Bhanu (2006)	2006
Gait History Image Liu and Zheng (2007)	2007
Gait Moment Image Ma et al. (2007)	2007
Frame Difference Energy Image Chen et al. (2009)	2009
Gait Entropy Image Bashir et al. (2009a)	2009
Motion Intensity Image + Motion Direction Image Bashir et al. (2009b)	2009
Active Energy Image Zhang et al. (2010)	2010
Chrono Gait Image Wang et al. (2010)	2010
Gait Flow Image Lam et al. (2011)	2011
Frame Difference History Image Lee et al. (2011)	2011
Poisson Random Walk Gait Energy Image Yogarajah et al. (2011)	2011
Depth Gradient Histogram Energy Image Hofmann et al. (2012)	2012

Table 2.1: Existing single compact 2D gait representations

3D approaches are commonly associated with increased computational demands due to setting up synchronised cameras. However equipment such as the Microsoft Kinect can extract a model-based 3D representation for gait recognition. Therefore this thesis uses 2D approaches for their suitability to image sequences captured from existing surveillance-type set ups.

2.1.2 Number of Images in the Gait Cycle Used to Represent Gait

The next debate considers the number of images in the gait cycle used to represent gait. The gait cycle, seen in Figure 2.1, is defined as the period taken between consecutive heel strikes from the same leg. Model-based approaches commonly use the complete gait cycle which requires higher computational demands. Alternatively, key frames [Collins

et al. (2002)] can be used which select a small number of discriminative images from fixed points in the gait cycle; note that this rejects a significant quantity of the gait cycle. Model-free approaches use an effective technique which condenses the gait cycle into a single compact 2D gait representation. These representations are advantageous due to the *i*) reduced computational demands and *ii*) natural robustness to noise and short term occlusion. Therefore single compact 2D gait representations inspire the research in this thesis. Examples of single compact 2D gait representations are shown in Table 2.1. Note that these representations are unique i.e. not based on re-engineering an existing single compact 2D gait representation; this is a separate topic which is discussed in Section 2.1.4 and also explains the lack of single compact 2D gait representations since 2012.

2.1.3 Gait Representations

This section discusses representations based on silhouettes, skeletons, contours and optical flow.

2.1.3.1 Silhouette gait representations

Silhouettes are a traditional technique to represent gait. These are simple, yet powerful descriptors due to the *i*) rejection of colour and texture cues which avoid bias to appearance given the inconsistency over time, *ii*) insensitivity to image quality, *iii*) reduced computational demands and *iv*) simple extraction via techniques such as background subtraction, Lidar, Time-of-flight or Microsoft Kinect.

The Gait Energy Image [Han and Bhanu (2006)] is a highly cited silhouette representation. The gait cycle image sequence is condensed into a single compact 2D gait representation by

$$GEI(x,y) = \frac{1}{N}\sum_{m=1}^{N} B_m(x,y) \tag{2.1}$$

where N is the number of silhouettes in the gait cycle, m is the silhouette number, x and y are the 2D spatial image coordinates and B is a silhouette. Note that (2.1) is analogous to a time-normalised Motion Energy Image [Bobick and Davis (2001)] action representation. The GEI, seen in Figure 2.2, shows static (pose) features and dynamic (motion) features which correspond to high and low pixel intensity values respectively. The time-

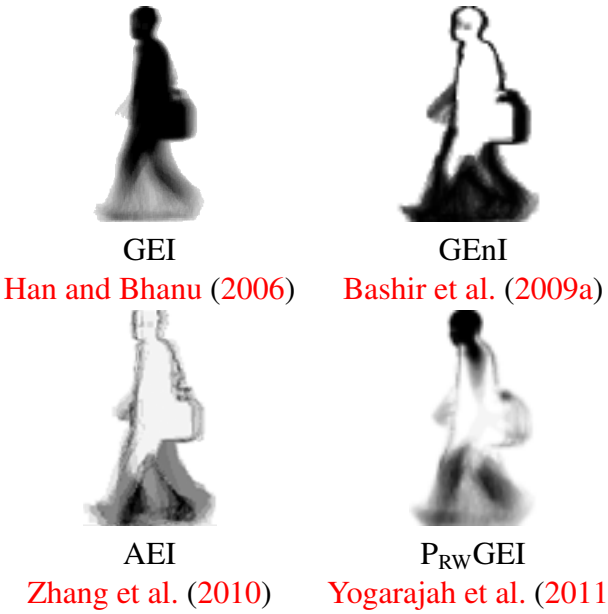


Figure 2.2: Silhouette gait representations are common and include the Gait Energy Image (GEI), Gait Entropy Image (GEnI), Active Energy Image (AEI), and Poisson Random Walk Gait Energy Image (P_{RW}GEI)

normalisation in (2.1) achieves natural robustness to noise and short term occlusion.

Alternative silhouette representations include the Gait Entropy Image (GEnI), Active Energy Image (AEI) and the Poisson Random Walk GEI (P_{RW}GEI). The GEnI [Bashir et al. (2009a)], seen in Figure 2.2, calculates the Shannon entropy (uncertainty) for each pixel by

$$H(x,y) = -\sum_{k=1}^{K} p_k(x,y) log_2 p_k(x,y) \tag{2.2}$$

where x and y are the 2D spatial image coordinates and p_k is the probability that x, y takes on the kth value. H is scaled and discretised for the GEnI to range between 0 and 255

$$GEnI(x,y) = \frac{(H(x,y) - H_{min}) * 255}{(H_{max} - Hmin)} \tag{2.3}$$

where $H_{min} = min(H(x,y))$ and $H_{max} = max(H(x,y))$. The GEnI shows discriminative motion features and rejects static features which are sensitive to covariate factors.

The AEI [Zhang et al. (2010)], seen in Figure 2.2, time-normalises the difference between consecutive images given

$$AEI(x,y) = \frac{1}{N} \sum_{t=0}^{N-1} D_t(x,y) \tag{2.4}$$

model-based skeleton Yoo and Nixon (2011) | model-free skeleton Whytock et al. (2014)

Figure 2.3: Unlike model-free skeletons, model-based skeletons are capable of maintaining individual leg positions during periods of self-occlusion

where D_t is the difference between consecutive silhouette pairs, x and y are the 2D spatial image coordinates and t is the number of silhouettes in the image sequence. The AEI is another representation which extracts discriminative motion features and rejects static features which are sensitive to covariate factors.

The P_{RW}GEI [Yogarajah et al. (2011)], seen in Figure 2.2, uses the Poisson Random Walk [Gorelick et al. (2004)] for covariate factor removal. To mitigate the appearance of covariate factors, the head and legs are extracted from a silhouette by using a threshold of

$$\psi = log(U(x,y) + ||\nabla U(x,y)||^2) \tag{2.5}$$

where U is the Poisson Random Walk, x and y are the 2D spatial image coordinates and ∇U is the gradient of the Poisson Random Walk. The P_{RW}GEI is given by

$$P_{RW}GEI(x,y) = \sum_{n=1}^{N} PRW_{sil}^{n}(x,y) \tag{2.6}$$

where PRW_{sil}^{n} is nth the Poisson Randon Walk silhouette in the image sequence, and x and y are the 2D spatial image coordinates.

2.1.3.2 Skeleton gait representations

Since the pioneering work of Blum (1967), skeletons have been used to represent shapes for numerous computer vision tasks. However skeleton representations are infrequently used for gait recognition as *a*) the walking action causes the body to self-occlude and *b*) silhouette quality has a direct impact on skeleton accuracy. Imperfect silhouette extraction causes boundary noise which manifests as additional unwanted skeleton spurs.

CGI
Wang et al. (2012)

Figure 2.4: The CGI uses coloured contours to encode temporal information in the single compact 2D representation

However with a degree of control, skeleton representations are compact and discriminative gait representations. Model-based approaches typically use the entire gait cycle to represent gait. Interestingly, there is a gap in knowledge relating to single compact 2D skeleton representations. This is a novel research area which this thesis will exploit.

Model-based approaches tend to construct skeletons which closely mimic human anatomy. Yoo and Nixon (2011) present a realistic 6-joint skeleton constructed from anthropometric data [Drillis and Contini (1966),Dempster and Gaughran (1967)]. The vertical positions of the neck, shoulder, waist, pelvis and ankle are connected to form the skeleton seen in Figure 2.3. This technique is less sensitive to boundary noise, and Figure 2.3 demonstrates that unique leg positions can be maintained during periods of self-occlusion. Note that the Microsoft Kinect can extract skeleton gait representations.

Model-free skeletons do not mimic human anatomy. As Figure 2.3 shows, model-free skeletons cannot differentiate between unique leg positions during self-occlusion. However the research in this thesis will demonstrate the success of model-free skeletons based on smooth distance functions generated from the Poisson equation. The smooth distance function reduces the sensitivity to boundary noise and yields a robust skeleton seen in Figure 2.3.

2.1.3.3 Contour gait representations

The contour of a silhouette is infrequently used to represent gait due to boundary noise sensitivity. An interesting contour gait representation is the Chrono-Gait Image (CGI) by Wang et al. (2012). The CGI addresses the lack of temporal information in a single compact 2D gait representation due to condensing a silhouette sequence. Temporal infor-

Bashir et al. (2009b)

Figure 2.5: The MII and MDIs are time-normalised to characterise the motion, and the motion in the four non-negative component directions respectively

mation is encoded by mapping the colour of each contour in the silhouette sequence by

$$CGI(x,y) = \frac{1}{p}\sum_{i=1}^{p} PG_i(x,y) \tag{2.7}$$

where p is the number of $\frac{1}{4}$ gait periods, x and y are the 2D spatial image coordinates and $PG_i(x,y) = \sum_{t=1}^{n_i} C_t(x,y)$ is the sum of the total n_i coloured contour images in the ith $\frac{1}{4}$ gait period. Figure 2.4 shows that the averaging in (2.7) alleviates boundary noise sensitivity.

2.1.3.4 Optical flow gait representations

Optical flow can extract a silhouette by analysing the motion between consecutive images in the gait cycle. Examples of optical flow gait representations include the *a*) Motion Intensity Image (MII) and Motion Direction Images (MDI) seen in Figure 2.5 and *b*) the Gait Flow Image (GFI) seen in Figure 2.6.

Bashir et al. (2009b) extract the RGB human figure from the image sequence and compute the optical flow field to yield a silhouette sequence. Instead of using the exact optical flow field, a Gaussian filter is applied to alleviate the sensitivity to noise. The MII, seen in Figure 2.5, time-normalises the optical flow field (the MII the equivalent to the GEI using optical flow). However, the MDIs in Figure 2.5 characterise motion by time-normalising the four non-negative component directions (similar to Efros et al. (2003)) denoted as $\mathbf{M}_x^+$, $\mathbf{M}_x^-$, $\mathbf{M}_y^+$ and $\mathbf{M}_y^-$. The pixel intensity values in Figure 2.5 correspond to the quantity of motion.

GFI
Lam et al. (2011)

Figure 2.6: The GFI uses optical flow to extract the motion in a silhouette sequence

Lam et al. (2011) calculate the optical flow field from consecutive silhouette images in a sequence (unlike the MII and MDIs which use segmented RGB figures). The GFI characterises the silhouette motion, seen in Figure 2.6, by calculating

$$GFI(x,y) = \frac{1}{N}\sum_{t=1}^{N-1} BF_{t,i}(x,y) \tag{2.8}$$

where N is the number of silhouettes in the sequence, t is the silhouette number, x and y are the 2D spatial image coordinates, i is the image sequence and BF is the optical flow field. The GFI is equivalent to the GEI using optical flow.

2.1.3.5 Discussion

The research in this thesis considers representations based on silhouettes and skeletons. Silhouette representations are used for the *a*) availability (standardised datasets freely provide silhouettes), *b*) low computational demand and *c*) insensitivity to image quality. Skeleton representations are chosen to exploit the knowledge gap relating to the novel combination of skeleton representations and single compact 2D gait representations.

Representations based on contours and optical flow are not used in this thesis. Contour representations encode the appearance of covariate factors. However covariate factor generalisation may be limited in contour representations due to the minimal contour appearance compared to silhouette representations. Gait recognition commonly uses optical flow to extract silhouettes, therefore the silhouette representation research could be equally applied to optical flow representations.

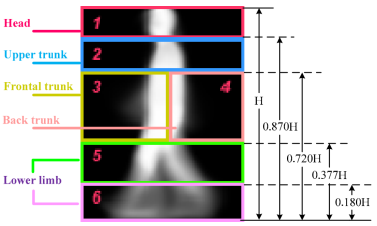
CASIA B dataset (124 test persons)

Novel Representations	average dataset performance (%)
Gait Energy Image Han and Bhanu (2006)	58.5
Motion Intensity and Direction Images Bashir et al. (2009b)	76.6
Gait Entropy Image Bashir et al. (2010)	74.1
Active Energy Image Zhang et al. (2010)	87.5
Poisson Randon Walk Gait Energy Image Yogarajah et al. (2011)	78.6
Chrono-Gait Image Wang et al. (2012)	58.2
average	72.3

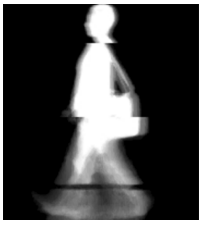
Re-engineering Existing Representations	average dataset performance (%)
M_G Bashir et al. (2008a)	90.5
Body segmentation Li et al. (2010)	85.2
Shifted Energy Image Huang and Boulgouris (2012)	78.3
Structural Gait Energy Image Li and Chen (2013)	87.6
average	85.4

Table 2.2: Average performance across all covariate factors in the CASIA B dataset when using new or re-engineering existing gait representations; the latter can be beneficial for covariate factor generalisation

M_G mask
Bashir et al. (2008a)



Body segmentation
Li et al. (2010)



SEI
Huang and Boulgouris (2012)

SGEI
Li and Chen (2013)

Figure 2.7: Examples of gait representations re-engineered from the GEI

2.1.4 Techniques to Improve Robustness

More recently, gait recognition approaches are divided between creating novel gait representations and re-engineering existing gait representations, i.e. developing an existing representation. Note that this division explains the lack of recent novel representations in Table 2.1. Consider the CASIA B dataset [Yu et al. (2006a),Zheng et al. (2011)] which is frequently used to validate gait recognition. Table 2.2 shows the difference between a high performing selection of the two distinct approaches. Novel gait representations achieve an average performance of 72.3%. However re-engineering existing gait representations achieve an average performance of 85.4%. Re-engineering an existing gait representation is successful as researchers can instead focus on developing novel covariate factor mitigation techniques.

The GEI uses a combination of space-normalisation and time-normalisation to condense a silhouette sequence into a single compact 2D gait representation. As such, the GEI is a common representation to re-engineer for the purpose of developing covariate factor mitigation techniques. The examples in Figure 2.7 include *a*) The M_G mask, *b*) body segmentation, *c*) Shifted Energy Image (SEI) and *d*) Structural Gait Energy Image (SGEI).

Bashir et al. (2008a) use the M_G mask, seen in Figure 2.7, as a covariate factor removal

mask which is given by

$$M_G(x,y) = \begin{cases} 1, & \text{if} \quad G_U(x,y) < \theta_1 \\ 0, & \text{if} \quad G_L(x,y) > \theta_2 \\ M_B(x,y), & \text{otherwise} \end{cases} \tag{2.9}$$

where M_B is a base mask, G_U represents the upper two thirds of the GEI, G_L represents the lower third of the GEI, x and y are the 2D spatial image coordinates, and θ_1 and θ_2 are pre-set thresholds. Covariate factor removal in (2.9) is achieved by using a threshold to remove static features (which are sensitive to covariate factors) and retain discriminative dynamic features.

Li et al. (2010) use anthropometric data [Drillis and Contini (1966),Dempster and Gaughran (1967)] to segment the body into six sections seen in Figure 2.7. The GEI is converted into a binary image by

$$B_x(x,y) = \begin{cases} 1, & \text{if} \quad T_S(x,y) > \theta \\ 0, & \text{if} \quad T_S(x,y) \leq \theta \end{cases} \tag{2.10}$$

where θ is a threshold ($0 < \theta < 255$), x and y are the 2D spatial image coordinates and T_S is the pixel-wise average of all the GEIs in the dataset. The pixel distribution in each section is characterised using (2.10) to determine if a covariate factor is present. If a covariate factor is detected, the section is removed from subsequent processing.

Huang and Boulgouris (2012) alleviate body rotations causes by covariate factors such as carrying a bag. Anthropometric data is used to segment the body into three sections, i.e. head, torso and legs. The centre of gravity for each section is horizontally aligned to form the SEI seen in Figure 2.7. Note that the SEI neglects natural body rotations which are discriminative.

Li and Chen (2013) use the SGEI, seen in Figure 2.7, which is equivalent to the GEI consisting of only the head and feet. The SGEI mitigates the appearance of covariate factors (assuming covariate factors do not affect the head and feet), however this neglects discriminative dynamic limb-based features.

2.1.4.1 Discussion

Constructing novel gait representations and re-engineering existing gait representations are both valid approaches for gait recognition. Researchers will continue to develop both approaches, and this thesis considers both approaches.

2.2 Action Recognition

A closely related topic to gait recognition, and also human motion analysis, is action recognition. Where gait recognition focusses on identifying a person by their unique walking manner, action recognition generalises over the unique walking manner to yield an action label in the form of a verb e.g. walk or run. Action recognition requires robustness to real world covariate factors, such as clothing and carrying a bag, which can alter the natural appearance (pose) and motion of an action. This section focusses on action representations and features; action classification techniques are not discussed in this section, however Poppe (2010) or Weinland et al. (2011) are recommended for further reading.

Action recognition lacks consistent terminology, where terms such as "action", "activity" and "event" have conflicting and often overlapping definitions [Moeslund et al. (2006),Laptev and Mori (2010)]. Standardised terminology is an open problem for action recognition, therefore the terminology from Moeslund et al. (2006) is used in this section:

- *Action primitives* are atomic movements e.g. right leg forward
- *Actions* are the combination of action primitives e.g. walking
- *Activities* [Aggarwal and Ryoo (2011)] are the combination of actions e.g. shopping

The simplest method to classify action representations is proposed by Poppe (2010) which defines global representations and local representations.

2.2.1 Global Representations

Global representations are constructed in a top-down manner. This requires person detection/segmentation to create a region of interest. These representations are effective due to

MEI MHI STV MHV
Bobick and Davis (2001) Blank et al. (2005) Weinland et al. (2006)]

Figure 2.8: Global representations for action recognition

the visibility of appearance (pose) and motion features. However global representations depend on accurate person detection/segmentation and are sensitive to covariate factors. Common global representations include silhouettes, skeletons, contours and optical flow.

Silhouettes are the simplest technique to visualise the human figure. Note that imperfect silhouette extraction causes boundary noise which can be alleviated with morphological operators. Bobick and Davis (2001) propose the *a*) Motion Energy Image (MEI) (2.11) and *b*) Motion History Image (MHI) (2.12). The MEI and MHI, seen in Figure 2.8, are single compact 2D representations which show where and how motion occurs respectively. The MEI E_τ is defined in (2.11) and the MHI H_τ is defined in (2.12) where

$$E_\tau(x,y,t) = \bigcup_{i=0}^{\tau-1} D(x,y,t-i) \tag{2.11}$$

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if} D(x,y,t)=1 \\ max(0, H_\tau(x,y,t-1)-1) & \text{otherwise} \end{cases} \tag{2.12}$$

where D is the silhouette image sequence and τ is the length of the image sequence. The Space-Time Volume [Blank et al. (2005)] and Motion History Volume [Weinland et al. (2006)] shown in Figure 2.8 are 3D equivalents of the MEI and MHI respectively.

Contours are derived from silhouettes. An example is the star skeleton [Chen et al. (2006)] where the 2D contour is unwrapped with respect to the figure centroid. This converts the 2D contour into a 1D signal where the maxima correspond to the head and limbs. When the maxima locations are connected to the figure centroid, the simplistic

Figure 2.9: Star skeleton [Chen et al. (2006)]

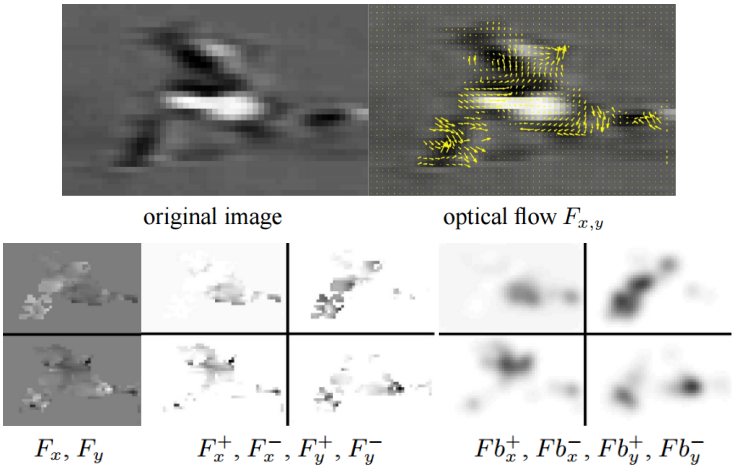


Figure 2.10: Blurry spatial patterns [Efros et al. (2003)]

skeleton in Figure 2.9 is formed. Note that the star skeleton cannot differentiate between unique legs during self-occlusion.

Optical flow can be used to extract the human figure by examining the motion between consecutive images. However dynamic backgrounds and camera motion can cause errors which can be alleviated via pre-processing. Efros et al. (2003) is a classical technique extracting optical flow vectors which are treated as blurry spatial patterns to form spatio-temporal motion descriptors seen in Figure 2.10.

2.2.2 Local Representations

Local representations are constructed in a bottom-up manner where detected space-time interest points (STIP) are described by local patches. These representations are effective as *i*) no person detection/segmentation is required, *ii*) the sensitivity to covariate factors is reduced and *iii*) there is a degree of invariance to background clutter and appearance. Note that local representations do not consider the human figure as a region of interest.

STIPs correspond to non-constant motion seen in Figure 2.11. Commonly used STIP detectors include Harris3D [Laptev and Lindeberg (2003)], Cuboid [Dollar et al. (2005)],

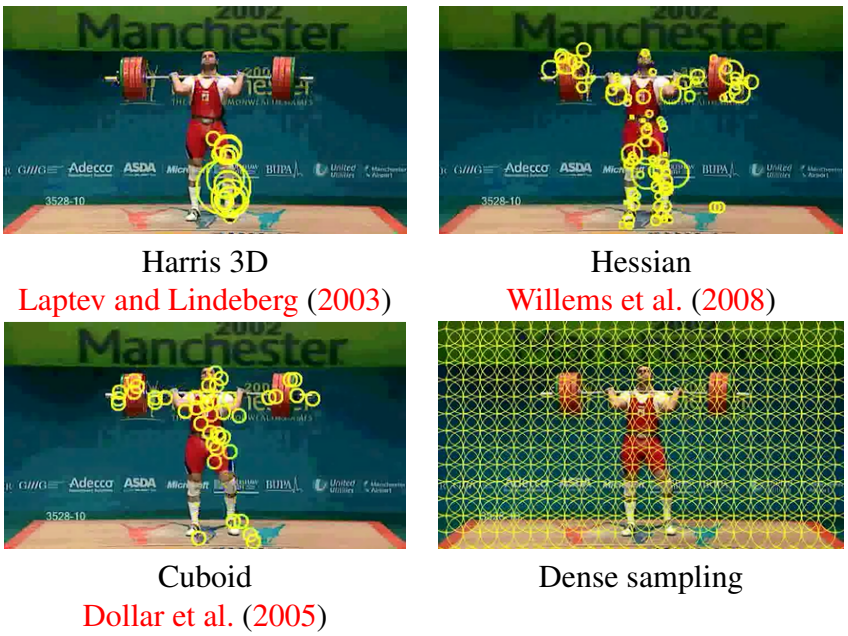


Figure 2.11: Space-time interest point detectors, adapted from Laptev and Mori (2010)

Hessian [Willems et al. (2008)] and dense sampling, while descriptors include Histograms of Oriented Gradients (HOG) and Histograms of Optical Flow [Laptev et al. (2008)], HOG3D [Kläser et al. (2008)] and extended Speeded Up Robust Features [Willems et al. (2008)]. Wang et al. (2009) provide a comprehensive evaluation of the aforementioned detectors and descriptors. Interesting conclusions suggest *i*) dense sampling (which generates a high quantity of features) achieves a higher performance which indicates a limitation of current STIP detectors and *ii*) Histograms of Oriented Gradients and Histograms of Optical Flow achieve a high performance which prompts further development. Note that STIP detectors often contain redundant features relating to the background. Therefore pre-processing is required to extract discriminative features relating to the human body.

2.2.3 Global/Local Grid-based Representations

Grid-based approaches divide an image into a spatial or temporal grid which alleviates the limitations of global representations and local representations. Local grid-based representations provide a degree of spatial (structural) information to local representations.

Histograms of Oriented Gradients [Dalal and Triggs (2005)] is a highly cited grid-based representation used for person detection [Dalal and Triggs (2005)], gait recogni-

Covariate Factors	
viewpoint	occlusion (full/partial/self)
injury	pregnancy
drunkenness	mood
weight	age
carrying a bag	anthropometrics
clothing (skirt, jackets)	shoes (flip flops, high heels)
speed	time

Table 2.3: Examples of covariate factors which affect the natural appearance and motion of human gait

tion [Sun et al. (2010),Hofmann et al. (2012)], action recognition [Laptev et al. (2008)] and gender recognition [Cao et al. (2008)]. Alternative global grid-based representations include *a*) Histograms of Oriented Optical Flow [Dalal et al. (2006),Laptev et al. (2008),Chaudhry et al. (2009a)] and *b*) Local Binary Patterns [Ojala et al. (2002)] which have been used for gait recognition [Kellokumpu et al. (2009)] and action recognition [Kellokumpu et al. (2008)].

2.2.4 Discussion

Global representations and global-grid based representations may be used for gait recognition. Note that local representations, specifically STIPs, have yet to be used for gait recognition (as gait is a full body movement, it is traditional for gait to be considered as a complete region of interest).

2.3 Covariate Factors

The term *covariate factor* is used by the gait recognition community to describe the challenges in gait recognition. Note that action recognition uses the term *challenges*. Existing research does not explicitly define covariate factors, therefore this thesis takes the opportunity to define a covariate factor.

Covariate factor

A covariate factor is a factor which affects the natural appearance and motion of human gait.

carrying a bag TUM GAID | occlusion TUM-IITKGP | viewpoint Weizmann | clothing CASIA B

Figure 2.12: Common covariate factors in existing datasets

Examples of gait recognition covariate factors can be seen in Table 2.3 (these covariate factors also apply to action recognition). Covariate factors can affect the appearance and motion of gait in combination or individually. For example, shoes such as flip flops can affect the natural motion of gait, clothing such as jackets can affect the appearance of gait, however carrying a bag can affect the appearance and motion of gait. The datasets discussed in Section 2.4 use a common set of covariate factors such as carrying a bag, occlusion, viewpoint and clothing seen in Figure 2.12. While covariate factors can be easily detected and identified by a human observer, mimicking the ability with computer vision is a challenging task. Gait recognition and action recognition require covariate factor robustness, and this is achieved by mitigating the effects of covariate factors.

There are other challenges which can indirectly affect the appearance of human gait and actions. For example, image noise, cluttered backgrounds and environment lighting can cause a human figure to be incompletely segmented. While pre-processing can alleviate the effects of incompletely segmented figures, constructing a robust gait representation is the foundation of gait recognition.

Discussion

Regardless of application, it is vital to extract discriminative features capable of generalising over as many covariate factors as possible to ensure robustness. This is vital for gait recognition especially, as while gait is unique, the differences between persons can be subtle. As a general rule, it is easier to generalise over single covariate factors compared to complex coupled covariate factors. Achieving a consistently high performance across covariate factors is an open problem and demonstrates the unlikelihood of developing a

single solution or parameter for gait recognition. This occurs as covariate factors uniquely affect the natural appearance and motion of human gait.

For gait recognition, research is divided between investigating the viewpoint covariate factor, or covariate factors such as carrying a bag, shoes, clothing and time (elapsed time between capture). This thesis considers the latter set of covariate factors in image sequences captured from a fixed side view. The side view is typical for gait recognition as it expresses the most discriminative limb motion. While capturing the side view is not always possible in real world unconstrained image sequences, viewpoint selection techniques [Rudoy and Zelnik-Manor (2012)] can be employed. Kale et al. (2003) or Seely et al. (2009) are recommended for techniques achieving robustness to viewpoint.

2.4 Datasets

Standardised datasets serve two vital purposes *a*) comparison against existing approaches and *b*) identification of potential limitations, which when identified and addressed, lead to increased robustness. Validation in multiple datasets is preferred to ensure the approach or parameters are not biased. Gait recognition and action recognition datasets are investigated in this section.

2.4.1 Dataset Requirements

Datasets should contain

- real world variation i.e. no choreography
- high person/action class numbers for inter-class and intra-class variation
- multiple image sequences for each person/action class
- separate and standard training sequences and test sequences
- real world single covariate factors and coupled covariate factors

2.4.2 Gait Recognition Datasets

Gait recognition has fewer standardised datasets (seen in Table 2.4) due to its infancy compared to action recognition. The majority of datasets freely provide silhouettes which has a two-fold advantage of *a*) enabling research to focus on the gait recognition problem as opposed to pre-processing problems e.g. silhouette extraction and *b*) creating a fair comparison against existing gait recognition approaches. In addition, these datasets have standard training sequences and test sequences.

The CASIA B dataset [Yu et al. (2006a),Zheng et al. (2011)] and TUM GAID dataset [Hofmann et al. (2012, 2013)], seen in Figure 2.13, are selected to validate the research in this thesis. Note that the CASIA B dataset captures persons walking from multiple viewpoints, however only the side views (90°) are used due to the amount of discriminative dynamic limb motion available. The CASIA B dataset is frequently used for validation which enables a thorough analysis against existing gait recognition approaches. Conversely, the TUM GAID dataset is a recent addition in Table 2.4 and includes complex coupled covariate factors. Overall, both datasets contain a variety of real world covariate factors and high person class numbers which are required to evaluate robustness.

Gait Recognition Datasets

Dataset/Year	Persons/Samples	Environment	Year	Covariate Factors
Soton Cunado et al. (1997)	10/40	Indoor	1997	
UCSD Little and Boyd (1998)	6/42	Outdoor	1998	time (minutes)
CMU MoBo Gross and Shi (2001)	25/600	Indoor (treadmill)	2001	viewpoint, walking speed, carrying condition, surface incline
Georgia Tech Johnson and Bobick (2001)	15/268	Outdoor	2001	time (months), viewpoint
	18/20	Magnetic tracker	2001	time (months)
HID-UMD 1 Kale et al. (2002)	25/100	Outdoor	2001	viewpoint, time
HID-UMD 2 Cuntoor et al. (2003)	55/220	Outdoor	2001	viewpoint, time
MIT Collins et al. (2002)	24/194	Indoor	2001	viewpoint, time
SOTON Small Nixon et al. (2001)	12/-	Indoor (chroma-key background)	2001	carrying condition, clothing, shoe, view
SOTON Large Nixon et al. (2001)	115/2128	Indoor, outdoor, treadmill	2001	view
HumanID Gait Challenge Sarkar et al. (2005)	122/1870	Outdoor	2002	viewpoint, surface, shoe carrying condition, time (months)
CASIA A Wang et al. (2003)	20/240	Outdoor	2001	viewpoint
CASIA B Yu et al. (2006a); Zheng et al. (2011)	124/13,640	Indoor	2005	viewpoint, clothing, carrying condition
CASIA C Tan et al. (2006)	153/1530	Outdoor (thermal camera)	2005	speed, carrying condition
OU-ISIR A Makihara et al. (2012)	34/408	Indoor (treadmill)	2007-2012	speed
OU-ISIR B Makihara et al. (2012)	68/1350	Indoor (treadmill)	2007-2012	clothing
OU-ISIR D Makihara et al. (2012)	185/370	Indoor (treadmill)	2007-2012	gait fluctuation
TUM-IITKGP Hofmann et al. (2011)	35/840	Indoor	2010	carrying condition, occlusion
SOTON Temporal Matovski et al. (2012)	25/2280	Indoor	2012	time, viewpoint
TUM GAID Hofmann et al. (2012, 2013)	305/3370	Indoor (regular + kinect)	2012	time (months), carrying condition, shoe

Table 2.4: Publicly available gait recognition datasets, adapted from Hofmann et al. (2013)

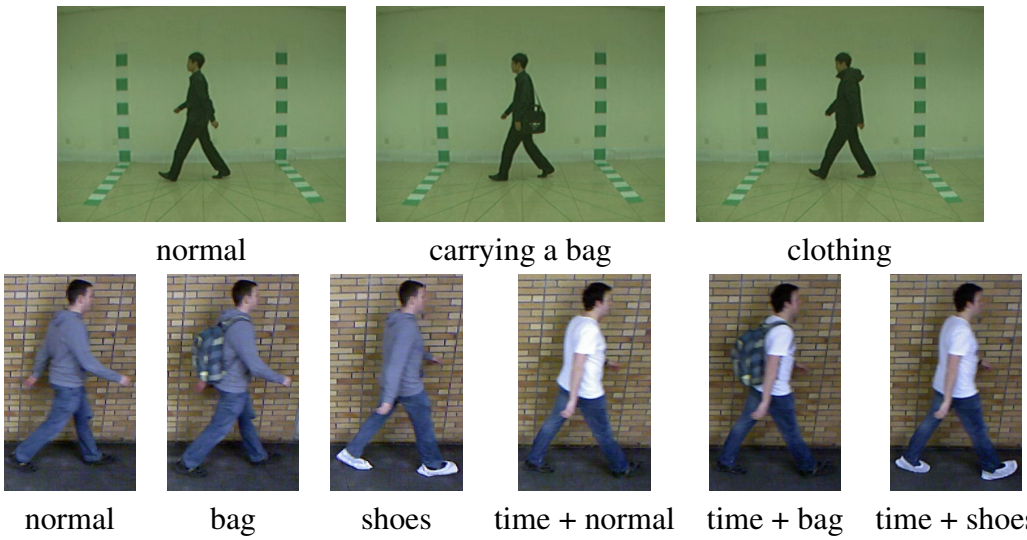


Figure 2.13: Images from the CASIA B dataset (top) and TUM GAID (bottom) dataset

TUM GAID Dataset Covariate Factors

Single Covariate Factors	Coupled Covariate Factors
normal (covariate factor free)	time and normal
carrying a bag	time and carrying a bag
shoes (clean room shoe covers)	time and shoes

Table 2.5: TUM GAID dataset covariate factors

2.4.2.1 CASIA B dataset

For nearly a decade the CASIA B dataset [Yu et al. (2006a),Zheng et al. (2011)], seen in Figure 2.13, has been used to validate gait recognition approaches. In an indoor environment, 124 persons have been captured under three covariate factors: *1*) normal i.e. covariate factor free, *2*) carrying a bag, which varies across the dataset e.g. handbags, rucksacks and *3*) clothing in the form of a bulky outdoor jacket which varies in length and shape. The dataset contains standard training sequences and test sequences. The training sequences use four normal (covariate factor free) sequences per person. The test sequences use two sequences per covariate factor per person.

2.4.2.2 TUM GAID dataset

The TUM GAID dataset [Hofmann et al. (2012, 2013)], seen in Figure 2.13, is captured in an indoor environment and is one of the largest datasets in Table 2.4. However its recency compared to the CASIA B dataset means fewer validation results exist. This

dataset contains single covariate factors and coupled covariate factors shown in Table 2.5. Six single covariate factors are used: *1*) normal i.e. covariate factor free, *2*) carrying a bag, which is consistent across the dataset and *3*) shoes, i.e. wearing clean room shoe covers. The three time-based coupled covariate factors are captured three months later. These covariate factors include clothing as an additional, yet hidden, covariate factor due to the change in weather season: *4*) time and normal, *5*) time and bag and *6*) time and shoes. The dataset contains standard training sequences and test sequences based on 305 persons and 155 persons respectively. The training sequences use four normal (covariate factor free) sequences per person. The test sequences use two sequences per covariate factor per person.

2.4.2.3 Silhouette quality comparison

The CASIA B dataset contains poorer quality silhouettes compared to the TUM GAID dataset due to the silhouette extraction technique. The TUM GAID dataset uses the Microsoft Kinect to extract depth information which yields relatively clean and intact silhouettes. Conversely, the CASIA B dataset uses background subtraction [Yu et al. (2006b)] which yields imperfect silhouettes containing extraneous noise causing missing heads or limbs. The difference in silhouette quality is advantageous as it is important for gait recognition to achieve a degree of robustness to silhouette quality.

2.4.3 Action Recognition Datasets

Publicly available action recognition datasets are shown in Table 2.6, where a recent survey of action recognition and activity recognition datasets can be found in Chaquet et al. (2013). It is uncommon for action recognition datasets to freely provide silhouettes. More recently, action recognition datasets are based on uncontrolled sequences from sources such as television [Niebles et al. (2010)], films [Laptev et al. (2008)] and online [Liu et al. (2009)].

Action Recognition Datasets

Name	Year	Name	Year
KTH Schuldt et al. (2004)	2004	HOLLYWOOD-2 Marszalek et al. (2009)	2009
ViSOR ViSOR (2011)	2005	MSR Action Yuan et al. (2009)	2009
WEIZMANN Actions Blank et al. (2005)	2005	UCF YouTube Liu et al. (2009)	2009
IXMAS Weinland et al. (2006, 2010)	2006/2010	URADL Messing et al. (2009)	2009
UCF Aerial UCF (b)	2007	MuHAVi Singh et al. (2010)	2010
HOLLYWOOD Laptev et al. (2008)	2008	Olympic Sports Niebles et al. (2010)	2010
UCF-ARG UCF (a)	2008	UCF50 Reddy and Shah (2013)	2010
UCF Sports Rodriguez et al. (2008)	2008	UT-Tower Chen and Aggarwal (2009)	2010
UIUC Action Tran and Sorokin (2008a)	2008	HMDB51 Kuehne et al. (2011)	2011
i3DPost Multi-view Gkalelis et al. (2009)	2009	VIRAT Oh et al. (2011)	2011

Table 2.6: Publicly available action recognition datasets, adapted from Chaquet et al. (2013)

2.4.4 Discussion

Action recognition datasets are divided into *a*) surveillance type sequences i.e. captured at a distance and therefore full body views and *b*) sequences from television, films and online sources and therefore partial body views. Note that some action recognition datasets may be applicable to gait recognition should unique person identification ground truth data be available. Gait recognition datasets use sequences captured from controlled environments. This causes gait recognition datasets to appear simplified compared to action recognition datasets. However with time, gait recognition approaches will achieve sufficient robustness to enable validation in unconstrained datasets similar to those for action recognition.

2.5 Summary and Motivations

This chapter discusses gait recognition (and the closely related topic of action recognition), covariate factors and standardised datasets. Section 2.1 demonstrates that gait recognition is an active and highly competitive research topic. The major debates in gait recognition include *1*) model-based, model-free and multi-information fusion approaches, *2*) the number of images in the gait cycle used to represent gait, *3*) gait representations and *4*) techniques to improve robustness. Current state-of-the-art gait recognition is based on *1*) model-free approaches, *2*) single compact 2D gait representations, *3*) silhouette representations and *4*) re-engineering existing representations with novel covariate factor mitigation techniques.

Covariate factor robustness remains the primary goal for researchers. The limitations of gait recognition approaches in Section 2.1 are not explicitly discussed as these are based on simplifying assumptions widely used by researchers. This thesis is motivated by addressing these limitations to achieve state-of-the-art gait recognition.

1. The degree and severity in which covariate factors affect the natural appearance and motion of gait is commonly underestimated. Despite covariate factors being located in a specific area, the effects can be seen across the body. For example, a rucksack can be found at the back of the torso and this affects the appearance of

the torso. However, the rucksack can cause a shifted centre of gravity, i.e. leaning, which affects the natural appearance and motion of gait. Therefore it is important to consider that the effects of covariate factors cannot be localised to a single area on the body.

2. The composition of covariate factors is commonly simplified. Covariate factors are static (more often than not) with respect to the human body. This means that researchers commonly consider covariate factors as static features. However this assumption neglects the fact that the natural motion of gait causes covariate factors to subsequently undergo motion. Therefore, covariate factors are composed of static and dynamic features.

Chapter 3

Gait Energy Image Described by Histograms of Oriented Gradients

This chapter is devoted to investigating the limitation of using Histograms of Oriented Gradients (HOG) parameters as a "black box". HOG is a feature descriptor first used in person detection where the parameters are tuned to describe the width of a limb in a single RGB human figure. However this chapter uses the grey-scale single compact 2D Gait Energy Image (GEI) to represent human gait/action. The combination of GEI and HOG is novel for action recognition. While the combination exists for gait recognition, HOG is fused with other features or validation is based on a small dataset. Therefore eight gradient schemes, and 100 combinations of cell size and bin sizes are evaluated to determine their contribution to robust action recognition and gait recognition. This chapter concludes that a high performance is achieved by tuning the HOG parameters.

Hypothesis

This chapter argues that to ensure robustness, HOG parameters (gradient scheme, cell size and bin size) must undergo re-evaluation when applied to applications other than person detection.

Publications

The results of this chapter have been presented at the British Machine Vision Conference Student Workshop [Whytock et al. (2012)] and the International Symposium on Visual Computing [Whytock et al. (2013a)].

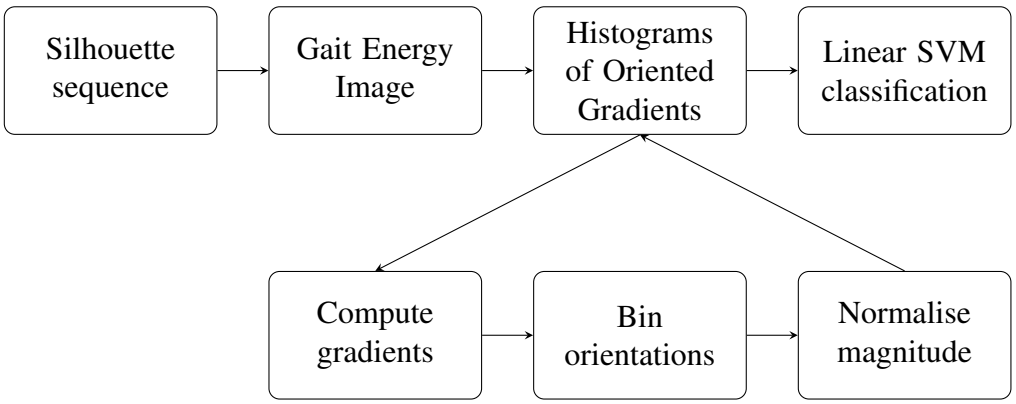


Figure 3.1: Proposed action recognition and gait recognition approach which converts a silhouette sequence into the Gait Energy Image representation and extracts features via Histograms of Oriented Gradients prior to Support Vector Machine (SVM) classification

This chapter combines the Gait Energy Image (GEI) and Histograms of Orientated Gradients (HOG). Human gait and human action are represented by the GEI and described by HOG. The GEI is a highly cited representation which condenses a silhouette sequence into a single compact 2D representation to extract static (pose) features and dynamic (motion) features. HOG is a benchmark feature descriptor which is an in-built function in Matlab R2014a. The core concept is based on describing an image via gradient distribution. The GEI and HOG combination, seen in Figure 3.1, is used for action recognition and gait recognition as these are closely related topics. Action recognition generalises over the action performance, while gait recognition focusses on the unique walking manner. The combination of GEI and HOG is novel for action recognition; however the combination exists for gait recognition. Sun et al. (2010) and Liu et al. (2012) show that the combination does not yield state-of-the-art gait recognition results unless feature fusion is applied or validation is based in a small dataset.

The HOG parameters defined by Dalal and Triggs (2005) are tuned for person detection and require re-evaluation due to multiple factors: *a*) HOG parameters are tuned to describe the width of a limb in a single RGB human figure, however the GEI is a greyscale representation which does not distinguish unique limbs and *b*) HOG was developed for person detection which is a two-class problem i.e. person versus not a person, however action recognition and gait recognition are multi-class problems where action classes and gait classes can be visually similar.

To this end, this chapter is devoted to evaluating HOG parameters (gradient scheme,

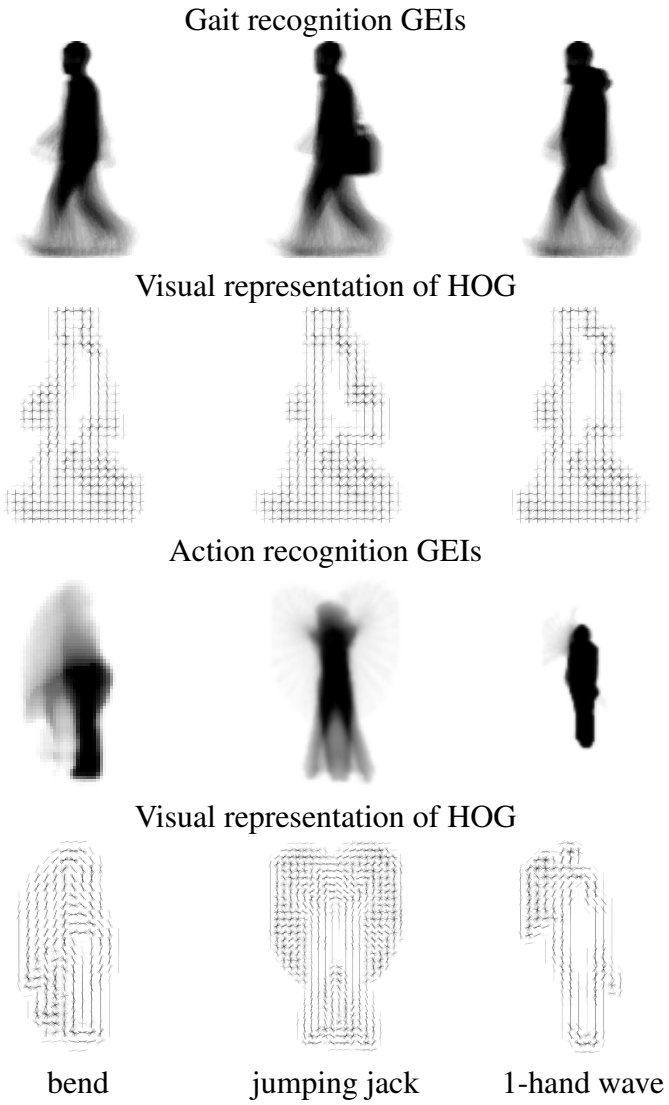


Figure 3.2: GEIs and corresponding HOG representations in the CASIA B dataset and Weizmann action dataset (centred gradient scheme, bin size $= 9$ and an excessively large cell size for illustrative purposes); notice how the distribution of pixel values and gradients vary between action recognition and gait recognition

cell size and bin size) for action recognition and gait recognition. In addition, a minor study on GEI body component contribution is performed for action recognition due to the varying distribution of pixel values seen in see Figure 3.2 (gait recognition is based on the walking action meaning GEIs display a regular distribution of pixel values). Further still, two classification schemes are evaluated for action recognition due to the loss of temporal information in the GEI.

3.1 Gait Energy Image

The Gait Energy Image (GEI) [Han and Bhanu (2006)] in Figure 3.2 represents human gait, but has also been used to represent actions [Lin et al. (2012)]. The GEI is founded on silhouettes which reject all colour and texture cues and provides a degree of invariance to gait appearance and action appearance given its inconsistency over time. The size-normalised silhouette sequence is condensed into the single compact 2D GEI given

$$GEI(x, y) = \frac{1}{N} \sum_{m=1}^{N} B_m(x, y) \tag{3.1}$$

where N is the number of silhouettes in the sequence, m is the silhouette number, x and y are the 2D spatial image coordinates and B is a silhouette. Note that the averaging caused by (3.1) causes a degree of temporal information loss.

Despite the compact nature, the GEI is a discriminative representation and is effective given the *i*) averaging aspect of (3.1) yields natural robustness to noise and short-term occlusion, *ii*) reduced computational demands and *iii*) extraction of static (pose) features and dynamic (motion) features which are differentiated by pixel intensity value. The low pixel intensity values correspond to limb motion which are discriminative dynamic features that are less sensitive to covariate factors. Conversely, high pixel intensity values correspond to static features from the head and torso which contain negligible motion; note that static features are sensitive to covariate factors.

3.2 Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) [Dalal and Triggs (2005)] is a highly cited feature descriptor employed in numerous computer vision applications. The core concept behind HOG is describing an image via gradient distribution. Dalal and Triggs (2005) proposed HOG for person detection, however HOG has now been used for gait recognition [Sun et al. (2010)], gender recognition [Cao et al. (2008)] and action recognition [Laptev et al. (2008)]. The limitation of HOG is the use as a "black box" which means parameters (gradient scheme, cell size and bin size) are not evaluated on the new application or image type.

3.2.1 Computing HOG Descriptors

A three stage process is required to extract HOG descriptors which is seen in Figure 3.1. Note that colour normalisation is not performed as the GEI is a grey-scale image which contains normalised pixel values given (3.1). *Stage 1: gradient computation*. *Stage 2: spatial/orientation binning* divides the image into cells where each pixel contributes a weighted vote for a gradient orientation-based histogram. *Stage 3: normalising descriptor blocks* is vital as gradient strength varies in the image. The square cells are grouped into larger spatial (square) blocks which overlap (half their area) meaning cells contribute multiple times to the final descriptor. This chapter uses *i*) square cells, *ii*) orientation bins spread evenly across 0° - 180° ("unsigned" gradient), *iii*) orientation voting based on gradient magnitude values and *iv*) block normalisation based on $L^2 - norm$.

3.2.1.1 HOG parameters of interest

There are two differences between the original person detection application [Dalal and Triggs (2005)] and the applications used in this chapter i.e. action recognition and gait recognition. Firstly, person detection is a two-class problem i.e. person versus not a person. However action recognition and gait recognition are multi-class problems where classes can be visually similar. Secondly, Dalal and Triggs (2005) use HOG on a single RGB image where parameter describe the width of a limb. However for action recognition and gait recognition, the GEI is a grey-scale single compact 2D representation where unique legs cannot be visualised. Therefore the optimum parameters for person detection may not be effective for action recognition and gait recognition.

This chapter focusses on the following HOG parameters: *a*) gradient scheme, *b*) cell size and *c*) bin size. While the cell size and bin size can be varied numerically, there are numerous gradient schemes available. Therefore this chapter uses a combination of simple traditional schemes recommended by Dalal and Triggs (2005) and sophisticated higher order, high accuracy gradient schemes such as those used in computational physics to model wave propagation phenomena.

3.2.2 Gradient Schemes

Dalal and Triggs (2005) trial 1D-point derivatives (uncentred, centred and cubic corrected), 2x2 diagonal masks and 3x3 Sobel masks. Overall, Dalal and Triggs (2005) conclude that the centred scheme achieves the highest performance for person detection where smoothing or large masks cause poorer performance. Therefore alongside the simple centred and Sobel masks trialled by Dalal and Triggs (2005), this chapter uses six additional higher accuracy gradient schemes: *1*) explicit Bickley [Bickley (1948)], *2*) explicit Scharr [Jähne et al. (1999)], *3*) implicit Bickley [Belyaev (2013)], *4*) implicit Scharr [Belyaev (2013)], *5*) Fourier-Padé-Galerkin [Belyaev (2011)] and *6*) Lele [Lele (1992a)]; these schemes increase with respect to computational demands respectively. The advantage of trialling a wide range of gradient schemes for action recognition and gait recognition is to analyse *i*) the gradient orientation and magnitude accuracy requirements and *ii*) the effect larger masks have on performance. In addition, this allows a comparison against person detection gradient scheme requirements whereby *i*) gradient orientation accuracy is of greater importance compared to gradient magnitude accuracy and *ii*) larger masks cause poorer performance. The eight gradient schemes are introduced individually before evaluating their gradient orientation and magnitude accuracy properties.

3.2.2.1 Traditional HOG gradient schemes

Given an image $I(x, y)$, first-order image derivatives are estimated via convolution with a kernel (mask)

$$D_x = \frac{1}{2h(1+2\alpha)} \left[\begin{array}{ccc} -\alpha & 0 & \alpha \\ -1 & 0 & 1 \\ -\alpha & 0 & \alpha \end{array} \right] \Bigg\} \begin{array}{l} \alpha = 0 \quad \text{(centred)}, \quad \alpha = \frac{1}{4} \quad \text{(explicit Bickley)} \\ \alpha = \frac{1}{2} \quad \text{(Sobel)}, \quad \alpha = \frac{3}{10} \quad \text{(explicit Scharr)} \end{array} \tag{3.2}$$

and its $\pi/2$-rotation, where h is the spacing between two neighbouring pixels and assume $h = 1$ for generality. Setting $\alpha = 0$ and $\alpha = 1/2$ in (3.2) yields the 1D centred and 3x3 Sobel masks respectively. Note that establishing an optimal α in (3.2) remains an open problem [Scharr et al. (1997),Weickert and Scharr (2002)]. Therefore (3.2) improves the centred scheme by incorporating smoothing in the orthogonal direction to compensate

for smoothing introduced by the centred scheme. This yields a more accurate gradient orientation estimation while simultaneously incorporating gradient magnitude smoothing. There is a low computational demand associated with the convolution in (3.2), therefore the centred and Sobel masks are termed explicit gradient schemes.

3.2.2.2 Higher accuracy explicit gradient schemes

Compared to the centred and Sobel masks, higher gradient orientation and magnitude accuracy is achieved given the observation by Bickley (1948)

$$D_x = \left(1 + \frac{h^2}{12}\Delta\right)\frac{\partial}{\partial x} + O(h^4) \quad \text{as} \quad h \to 0 \tag{3.3}$$

where Δ is the Laplacian and is rotationally invariant. Setting $\alpha = 1/4$ in (3.2) yields the explicit Bickley scheme, while setting $\alpha = 3/10$ in (3.2) yields the explicit Scharr scheme. The explicit Bickley and explicit Scharr schemes are similarly termed explicit gradient schemes due to the low computation demands required to implement (3.3).

3.2.2.3 Higher accuracy implicit gradient schemes

By combining differencing and smoothing in a slightly different manner, increased gradient orientation and magnitude accuracy can be achieved. Finite differencing is combined with inverted smoothing (sharpening) to yield implicit/compact finite differences which are standard tools for applications such as computational physics [Lele (1992b)] e.g. modelling wave propagation phenomena.

Therefore, the implicit Bickley scheme corresponding to (3.2) with $\alpha = 1/4$ is

$$\frac{1}{6}\left[I'_x(x-h,y) + 4I'_x(x,y) + I'_x(x+h,y)\right] \quad \approx \quad \frac{I(x+h,y) - I(x-h,y)}{2h} \tag{3.4}$$

where the x- and y-derivatives are similarly estimated. Notice the smoothing introduced by the central difference scheme in the left-hand side of (3.4) is compensated by incorporating smoothing (averaging) on the right-hand side. The implicit Bickley scheme is achieved by setting $\alpha = 1/4$ in (3.4) and corresponds to the 4th-order Padé approximation [Belyaev (2011)]. Similarly, the implicit Scharr scheme is achieved by setting $\alpha = 3/10$

in (3.4). The implicit Bickley and implicit Scharr schemes are termed implicit gradient schemes as (3.4) requires solving a tridiagonal system of linear equations.

The Fourier-Padé-Galerkin and Lele schemes achieve superior accuracy using

$$\beta f'_{i-2}+\alpha f'_{i-1}+f'_i+\alpha f'_{i+1}+\beta f'_{i+2} = c\,\frac{f_{i+3}-f_{i-3}}{6}+b\,\frac{f_{i+2}-f_{i-2}}{4}+a\,\frac{f_{i+1}-f_{i-1}}{2} \tag{3.5}$$

where α, β, a, b and c are defined by each scheme and are optimised to ensure smoothing applied to the right-hand side of (3.5) is compensated by averaging the derivatives on the left-hand side. Therefore, the Lele [Lele (1992a)] and Fourier-Padé-Galerkin [Belyaev (2011)] schemes use the following coefficients

$$\left.\begin{aligned} \alpha = 0.5771439, \quad \beta = 0.0896406 \\ a = 1.302566, \quad b = 0.99355, \quad c = 0.03750245 \end{aligned}\right\} \quad \text{Lele scheme} \tag{3.6}$$

$$\left.\alpha = \frac{3}{5}, \quad \beta = \frac{21}{200}, \quad a = \frac{63}{50}, \quad b = \frac{219}{200}, \quad c = \frac{7}{125}\right\} \quad \text{Fourier-Padé-Galerkin scheme} \tag{3.7}$$

The Fourier-Padé-Galerkin and Lele schemes require solving a system of linear equations and are therefore termed implicit gradient schemes. Note that these schemes are computationally equivalent to convolution with a 5x5 mask.

3.2.2.4 Gradient scheme comparison

A circular pattern sinusoidal grating is used as a visual technique to demonstrate the gradient magnitude and orientation accuracy properties of each gradient scheme. Note that the frequency response is an alternative method, however this is based on 1D cases which do not reveal the gradient magnitude and orientation properties. The sinusoidal grating seen in Figure 3.3 is given by

$$\sin(x^2+y^2) \tag{3.8}$$

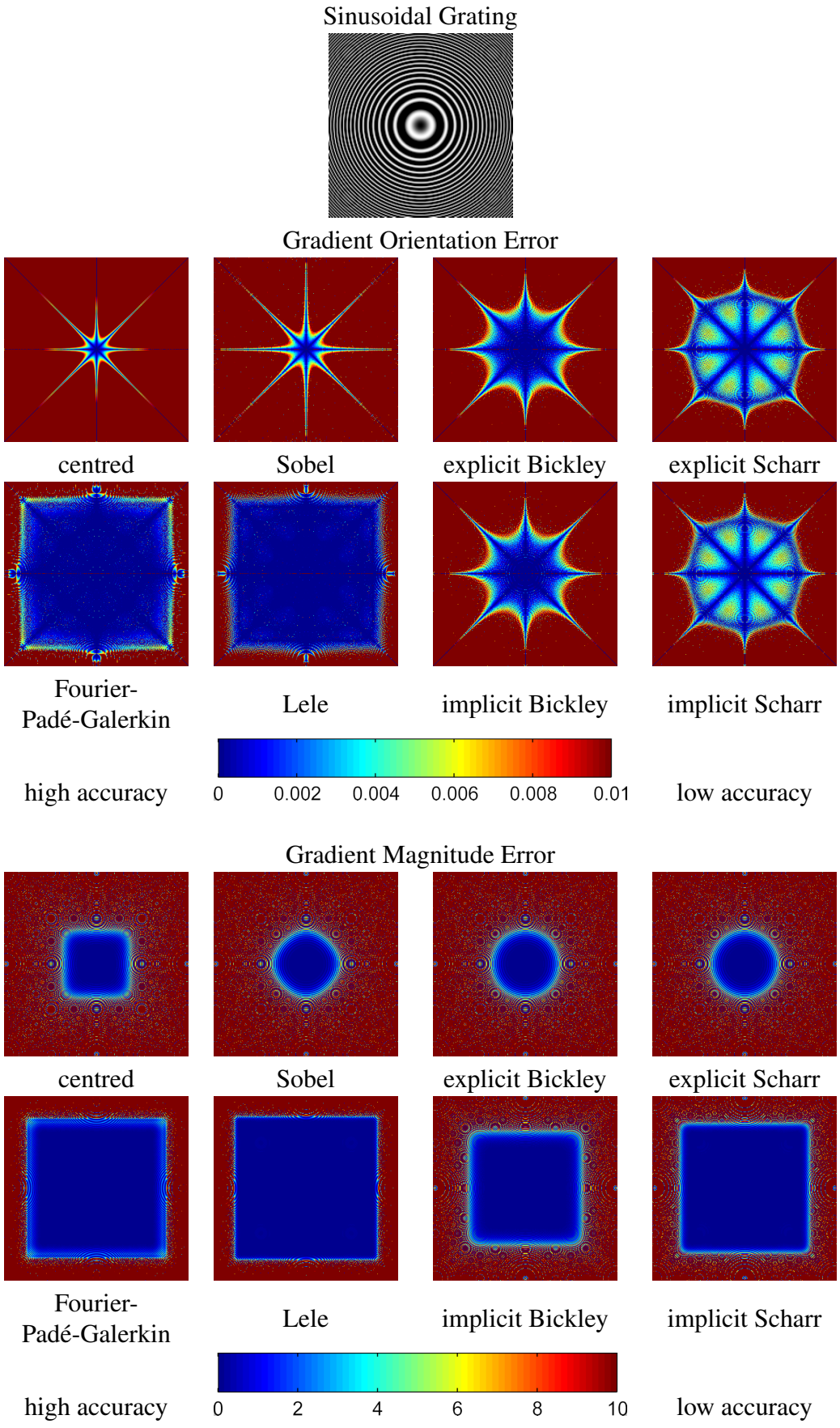


Figure 3.3: A sinusoidal grating is used to visualise the gradient magnitude and orientation accuracy for each gradient scheme

and contains low-range frequencies at the centre which radiate out to medium-range frequencies. This is achieved by numerical experiments with varying frequencies, i.e. 0 to 0.5 cycles/pixel (the Nyquist frequency). As the GEI does not contain texture cues, high-range frequencies are not included in the sinusoidal grating.

Gradient orientation. While the centred and Sobel masks achieve low-range frequency accuracy, the Fourier-Padé-Galerkin and Lele schemes achieve low-range and medium-range frequency accuracy. A compromise with respect to accuracy is met by using the explicit Bickley and explicit Scharr schemes. Note that the implicit Bickley and implicit Scharr schemes achieve the same accuracy as their explicit counterparts with no additional computational demands.

Gradient magnitude. The explicit gradient schemes achieve low-range frequency accuracy, while the Fourier-Padé-Galerkin and Lele schemes achieve low-range and medium-range accuracy. The implicit Bickley and implicit Scharr schemes achieve higher gradient magnitude accuracy compared to their explicit counterparts with no additional computational demands.

3.3 Application: Action Recognition

The combination of GEI and HOG is novel for action recognition, therefore a quantitative and qualitative evaluation is required for the gradient scheme, cell size and bin size HOG parameters. Action recognition disregards the unique action performance to extract an action verb e.g. walk. The GEIs seen in Figure 3.4 show an inconsistent distribution of static and dynamic features (GEIs in gait recognition show a repetitive distribution of static and dynamic features due to only considering the walking action). Therefore this section analyses the contribution of the body to performance. This is achieved by considering the full body, upper body and lower body in the GEIs. In addition, as a degree of temporal information is lost in (3.1), a separate analysis is performed to determine how classification schemes which include temporal information can boost performance.

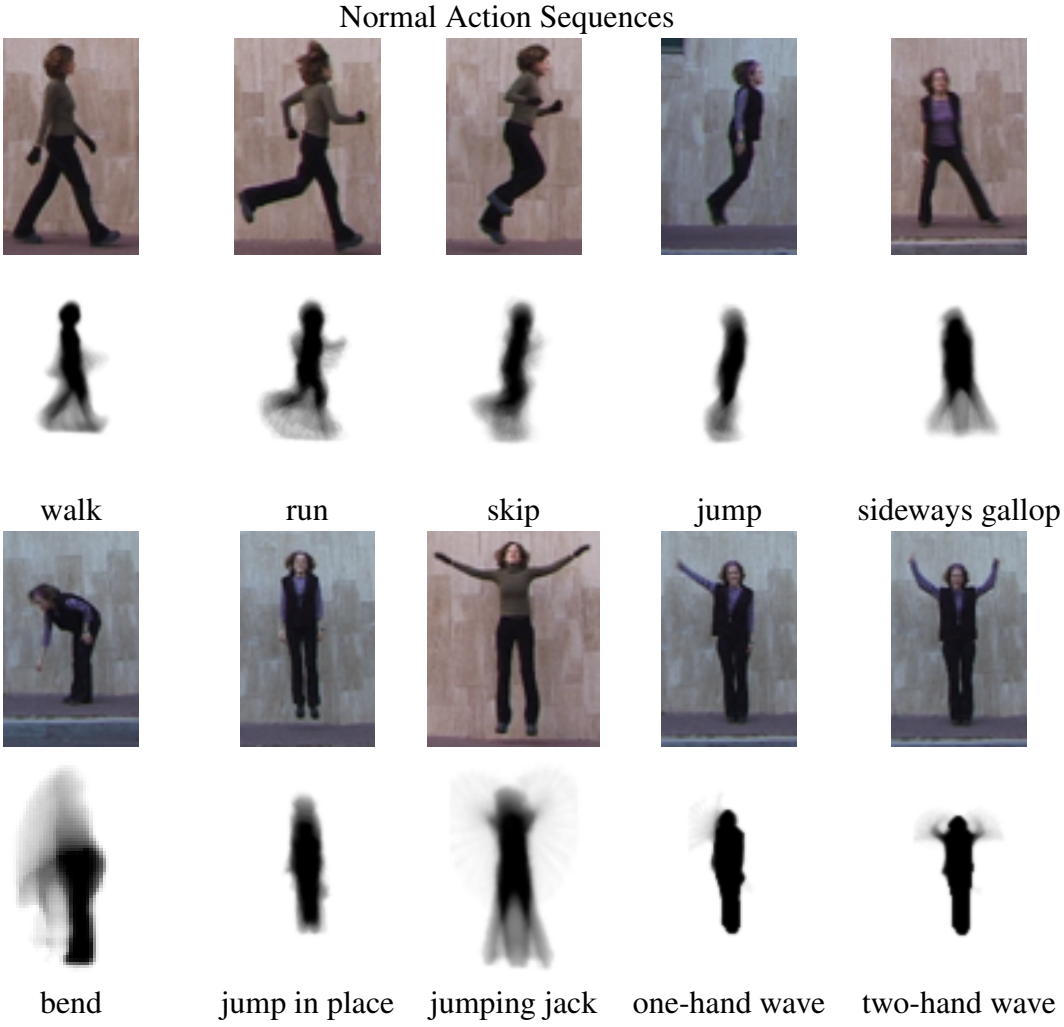


Figure 3.4: GEIs from the Weizmann Action dataset and their corresponding original RGB images; notice that these actions are covariate factor free

Figure 3.5: Weizmann Action dataset deformation (appearance-based and motion-based covariate factors) GEIs with their corresponding original RGB images

3.3.1 GEIs Representing Action

The Weizmann Action dataset [Blank et al. (2005)] contains three sets of image sequences: *a*) normal (covariate factor), *b*) deformation (appearance-based and motion-based covariate factors e.g. clothing and carrying a bag) and *c*) viewpoint.

The normal (covariate factor free) action sequences are seen in Figure 3.4 and show a unique distribution of static and dynamic features. Notice that the views chosen to represent the actions in Figure 3.4 express the highest quantity of dynamic features. The deformation action sequences are based on the walking action and are seen in Figure 3.5. These action sequences are captured from the side view as this expresses the highest quantity of dynamic features. The viewpoint sequences are also based on the walking

Figure 3.6: Weizmann Action dataset GEIs affected by view-based covariate factors with their corresponding original RGB images

action. Figure 3.6 shows that views deviating from 0° (side view) show progressively fewer dynamic features.

The GEI is effective as (3.1) achieves natural robustness to short term occlusions. This is demonstrated in Figure 3.5 where the "pole" from the "occluded by a pole" sequence causes a short term occlusion of the person. Notice how the GEI retains the dynamic features relating to lower limb motion. However the GEI cannot overcome long-term occlusions seen in the "occluded feet" sequence. In this case the GEI shows minimal dynamic features from lower limb motion.

3.3.2 Experimental Procedure: Action Recognition

3.3.2.1 GEI construction

The GEI requires space-normalisation prior to (3.1). This is achieved by horizontally aligning the centroid of the top 10% of the silhouette (head) as a reference point. This process is similar to Hofmann et al. (2011).

3.3.2.2 Dataset

The classical Weizmann Action dataset [Blank et al. (2005)], presented in Figure 3.4, Figure 3.5 and Figure 3.6, is used for validation due to *i*) the extensive use in action recognition validation and *ii*) the image sequences are pre-processed meaning silhouettes are freely available and space-normalised ready for (3.1). Nine persons (of mixed gender) perform ten actions: *i*) run, *ii*) walk, *iii*) skip, *iv*) jumping jack *v*) jump in place, *vi*) jump, *vii*) sideways gallop *viii*) one-hand wave, *ix*) two-hand wave and *x*) bend. The Weizmann Action dataset divides normal (covariate factor free) action sequences and robustness (deformation and viewpoint covariate factor) action sequences. Note that the robustness sequences are based on the walking action only. The robustness sequences are further divided into deformation (appearance-based and motion-based covariate factors) and viewpoint sequences. Deformation sequences include *i*) walk with a dog, *ii*) swinging a bag, *iii*) walk in a skirt, *iv*) occluded feet, *v*) occluded by a "pole", *vi*) moonwalk, *vii*) limp walk, *viii*) walk with knees up, *ix*) walk with a briefcase and *x*) normal walk. The viewpoint sequences are captured from 0° to 81° in increments of 9° which reflect the side view heading towards a frontal view.

3.3.2.3 Classification

Following the traditional HOG implementation by Dalal and Triggs (2005), this chapter uses Linear Support Vector Machine (SVM) [Cortes and Vapnik (1995)] classification. Weizmann Action dataset standards require leave-one-sequence-out cross validation. The SVM results are compared against the ground truth in a confusion matrix. The average of the diagonal in the confusion matrix yields the correct classification.

3.3.2.4 Unique experiments

There are three experiments to be performed: *i*) HOG parameter evaluation (gradient scheme, cell size and bin size), *ii*) body component contribution and *iii*) classification scheme.

HOG parameters

Eight gradient schemes are evaluated: *1*) centred, *2*) Sobel, *3*) explicit Bickley, *4*) explicit Scharr, *5*) implicit Bickley, *6*) implicit Scharr, *7*) Fourier-Padé-Galerikin and *8*) Lele. One hundred combinations of cell size and bin size are evaluated (note that a large cell size and bin size yields high dimensionality feature vectors). Cell sizes c range from $\{c = 1$ in steps of 1 to $c = 10\}$ and bin sizes b range from $\{b = 3$ in steps of 3 to $b = 30\}$.

Body component contributions

Action GEIs vary with respect to static feature and dynamic feature distribution. Therefore the body is considered as a whole (full body GEI) and split into the upper body GEI and lower body GEI using anthropometric data [Drillis and Contini (1966),Dempster and Gaughran (1967)].

Classification schemes

Two classification schemes are proposed to analyse the effect of including temporal information (somewhat lost in (3.1)) at the classification stage. The first scheme classifies all ten actions during SVM classification. The second scheme includes a degree of temporal information by splitting actions into static and dynamic action classes based on a manual threshold of global translation, i.e. those which remain in one position over time and those which move globally respectively. There is an equal split of five dynamic actions *i*) run, *ii*) walk, *iii*) skip, *iv*) jump and *v*) sideways gallop, and five static actions *i*) bend, *ii*) jump in place, *iii*) jumping jack, *iv*) one-hand wave and *v*) two-hand wave.

Body Component Contribution (normal action sequences)		
Full body GEI (%)	Lower body GEI (%)	Upper body GEI (%)
97.4	87.5	84.7

Table 3.1: Weizmann Action dataset body component contribution averaged across gradient scheme and classification scheme

3.3.3 Results and Discussion

The results are evaluated separately for normal (covariate factor free) action sequences and robustness (covariate factor) sequences in the Weizmann Action dataset. The normal action sequences are used to demonstrate the performance based on *i*) body component contribution, *ii*) gradient scheme *iii*) HOG cell size and bin size and *iv*) classification scheme. The robustness sequences are used to demonstrate how robustness varies when using *i*) HOG parameters achieving the highest normal action sequence performance and *ii*) HOG parameters achieving the highest robustness sequence performance.

3.3.3.1 Normal action sequence evaluation

Normal action sequences are used to evaluate the performance for each *i*) body component contribution in Table 3.1, *ii*) gradient scheme in Table 3.2, *iii*) HOG cell size and bin size in Figure 3.7, Figure 3.8 and Figure 3.9 and *iv*) classification scheme in Table 3.3.

Body component contribution

The distribution of static features and dynamic features vary with the action performed and body segment under consideration, i.e. static action GEIs are dominated by upper limb motion while dynamic action GEIs contain a mixture of upper and/or lower limb motion. The contribution of the body to performance is seen in Table 3.1 where the performance is averaged across the gradient schemes and classification schemes. The full body GEIs achieve a high performance due to the quantity of discriminative features available. As the actions contain a varied distribution of static and dynamic features, the lower body and upper body rank closely second and third respectively. These results demonstrate that the upper body and lower body, and therefore static and dynamic features, are discriminative for action recognition. Note that gait recognition [Bashir et al. (2008b),Martín-Félez and

Gradient Scheme Performance (normal action sequences)

Explicit schemes (%)		Implicit schemes (%)	
centred	92.2	implicit Bickley	90.0
Sobel	89.8	implicit Scharr	90.6
explicit Bickley	92.6	Fourier-Padé-Galerkin	91.3
explicit Scharr	92.6	Lele	91.9

Table 3.2: Weizmann Action dataset gradient scheme performance averaged across body component contribution and classification scheme

Xiang (2012)] and gait-based gender recognition [Yu et al. (2009)] demonstrate a greater difference between the upper body and lower body contributions. Lower limb dynamic features are more discriminative for gait recognition, while upper body static features are more discriminative for gait-based gender recognition.

Gradient scheme

The gradient scheme results seen in Table 3.2 are averaged across body component contribution and classification scheme. Compared to the centred scheme, a Wilcoxon test ($p > .05$, two tailed test) indicates that action recognition results are not significantly affected by the gradient scheme. In addition, a Wilcoxon test ($p > .05$, two tailed test) indicates that action recognition results are not significantly affected by the type (explicit/implicit) of gradient scheme. These results suggest that *a*) similar to person detection, action recognition requires higher gradient orientation accuracy compared to gradient magnitude accuracy and *b*) contrary to person detection, larger 3x3 gradient scheme masks can increase performance.

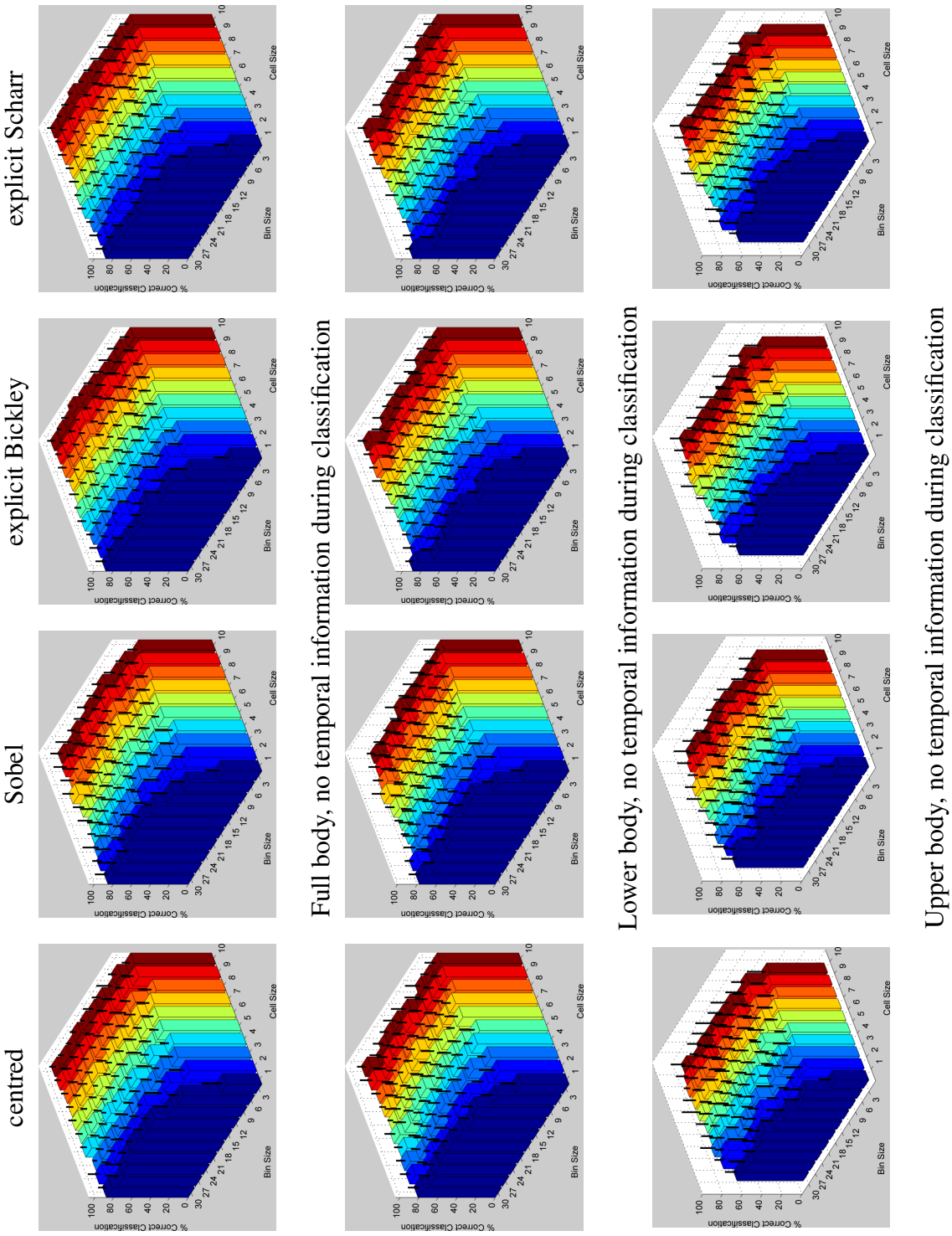


Figure 3.7: Weizmann Action dataset normal action (10 classes) sequence performance with respect to HOG cell size and bin size, and body component contribution (full, lower and upper body) for each gradient scheme when no temporal information is included during classification; the error bars are shown in black

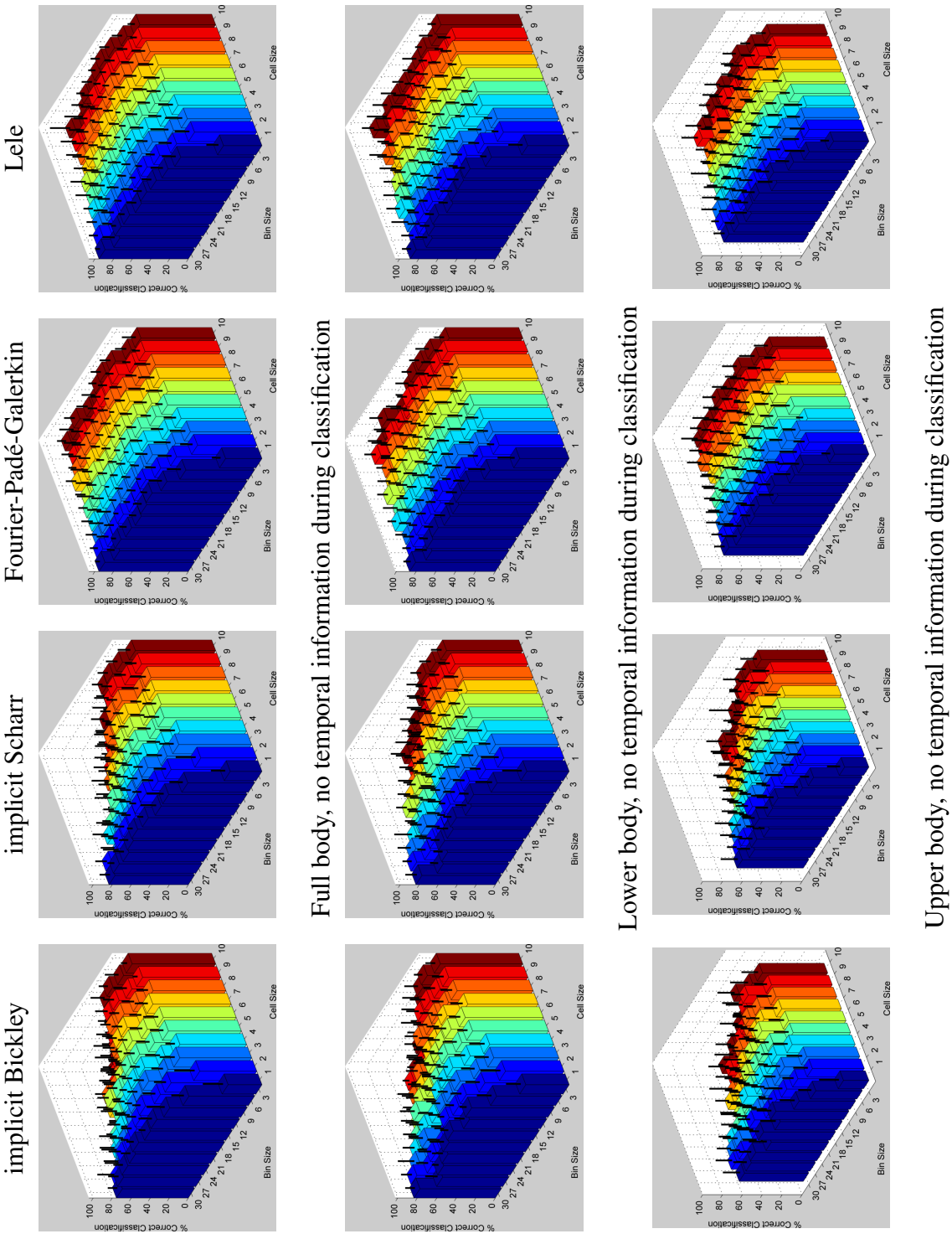


Figure 3.7 (Continued): Weizmann Action dataset normal action (10 classes) sequence performance with respect to HOG cell size and bin size, and body component contribution (full, lower and upper body) for each gradient scheme when no temporal information is included during classification; the error bars are shown in black

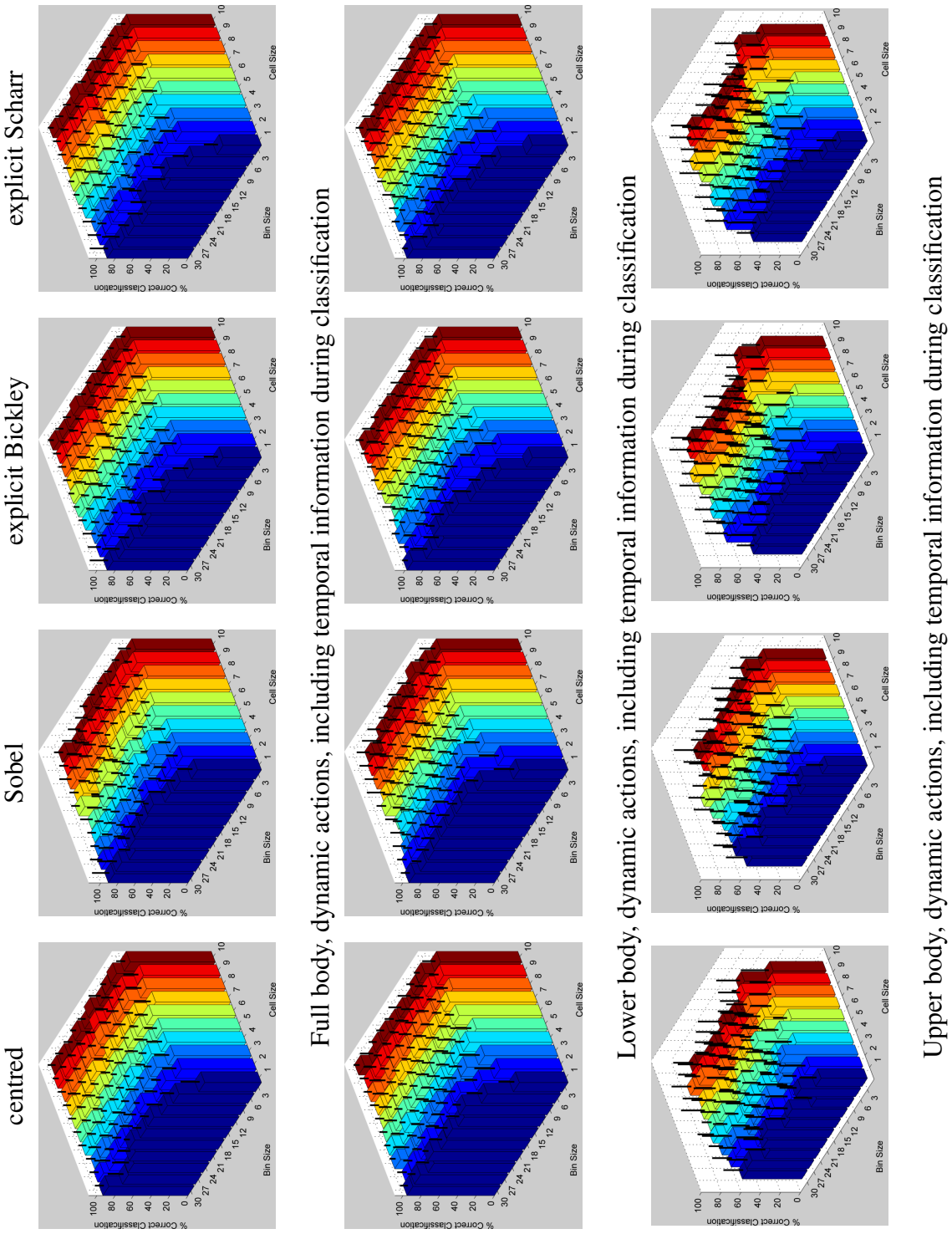


Figure 3.8: Weizmann Action dataset normal action (5 classes) sequence performance with respect to HOG cell size and bin size, and body component contribution (full, lower and upper body) for each gradient scheme when incorporating temporal information by classifying dynamic actions only; the error bars are shown in black

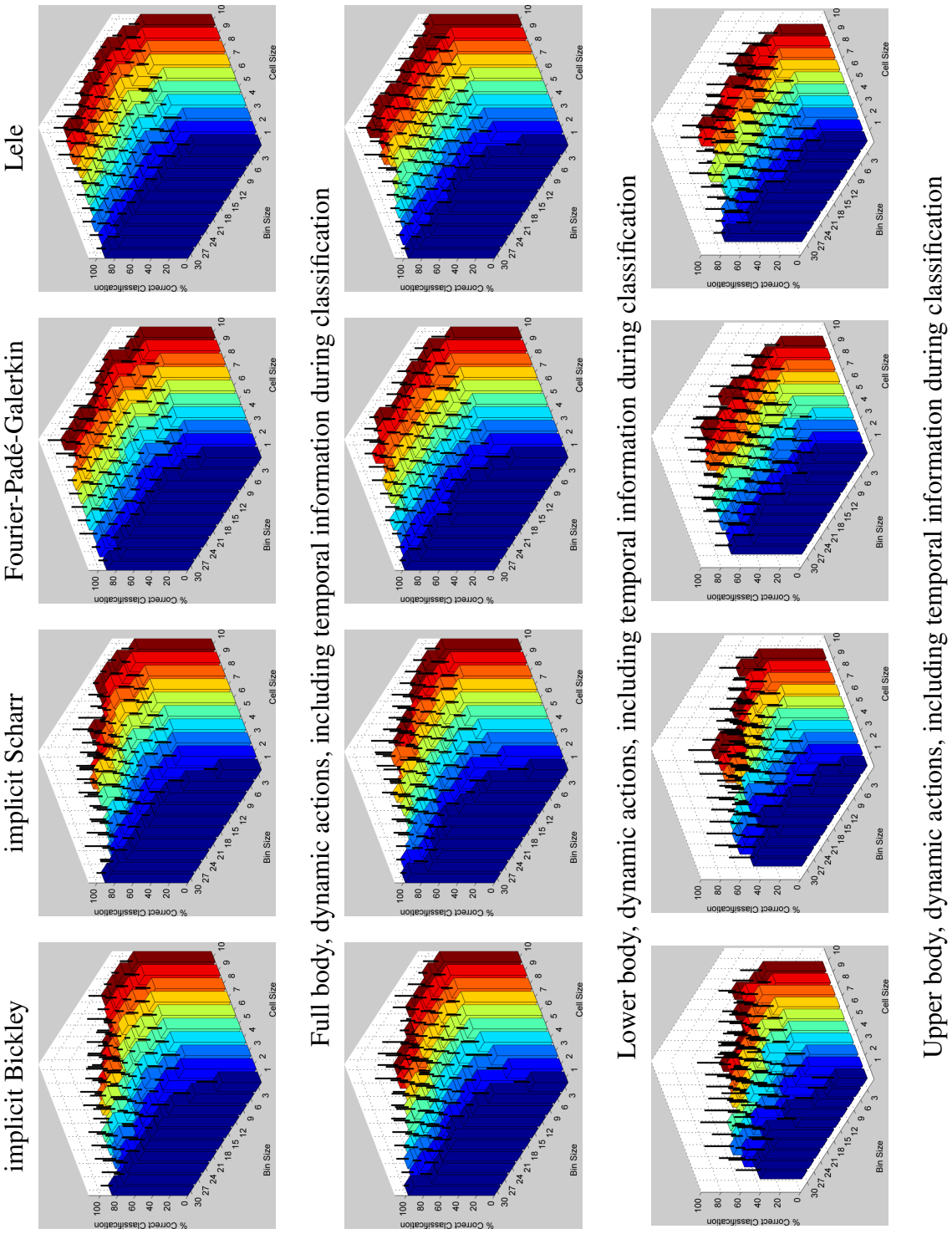


Figure 3.8 (Continued): Weizmann Action dataset normal action (5 classes) sequence performance with respect to HOG cell size and bin size, and body component contribution (full, lower and upper body) for each gradient scheme when incorporating temporal information by classifying dynamic actions only; the error bars are shown in black

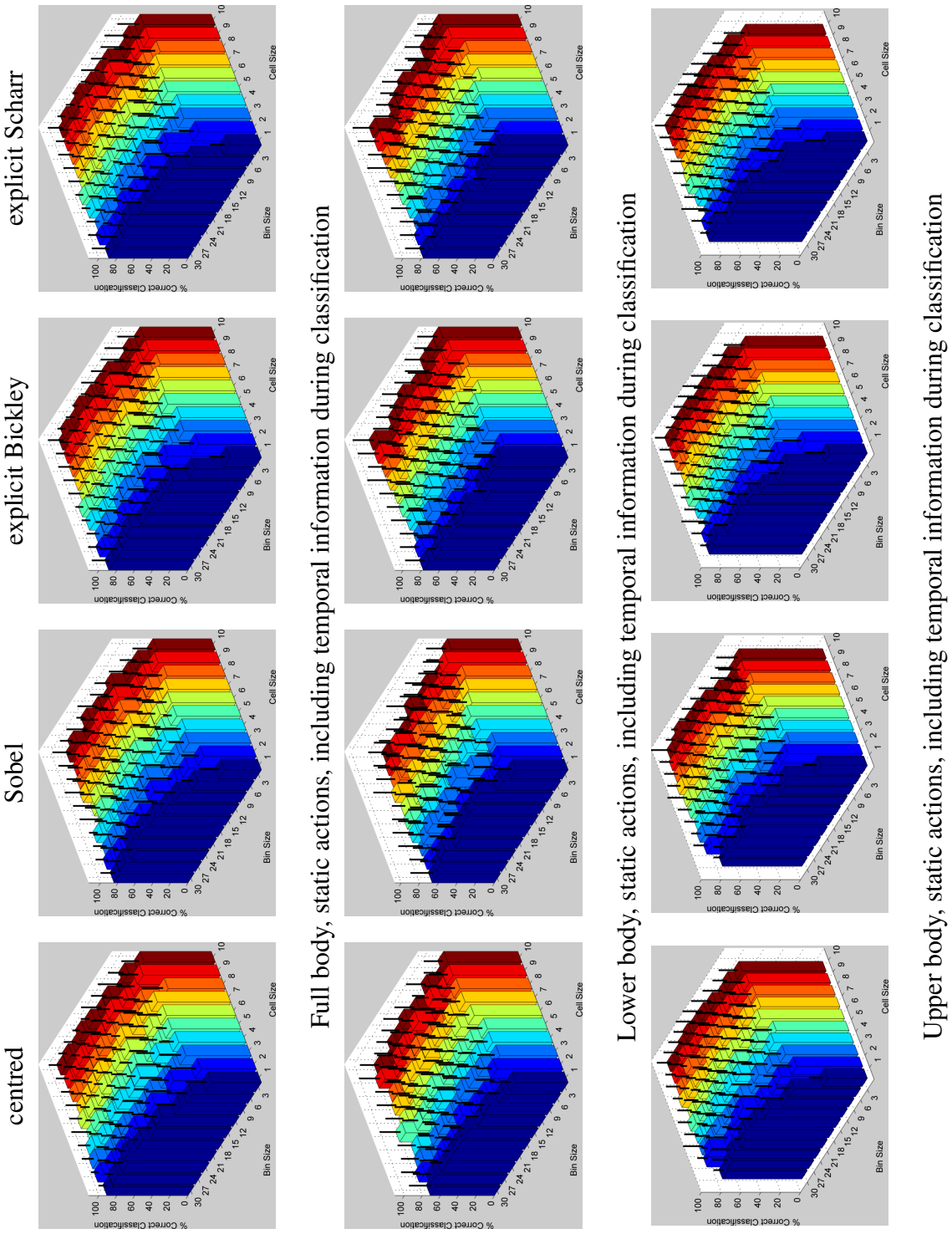


Figure 3.9: Weizmann Action dataset normal action (5 classes) sequence performance with respect to HOG cell size and bin size, and body component contribution (full, lower and upper body) for each gradient scheme when incorporating temporal information by classifying static actions only; the error bars are shown in black

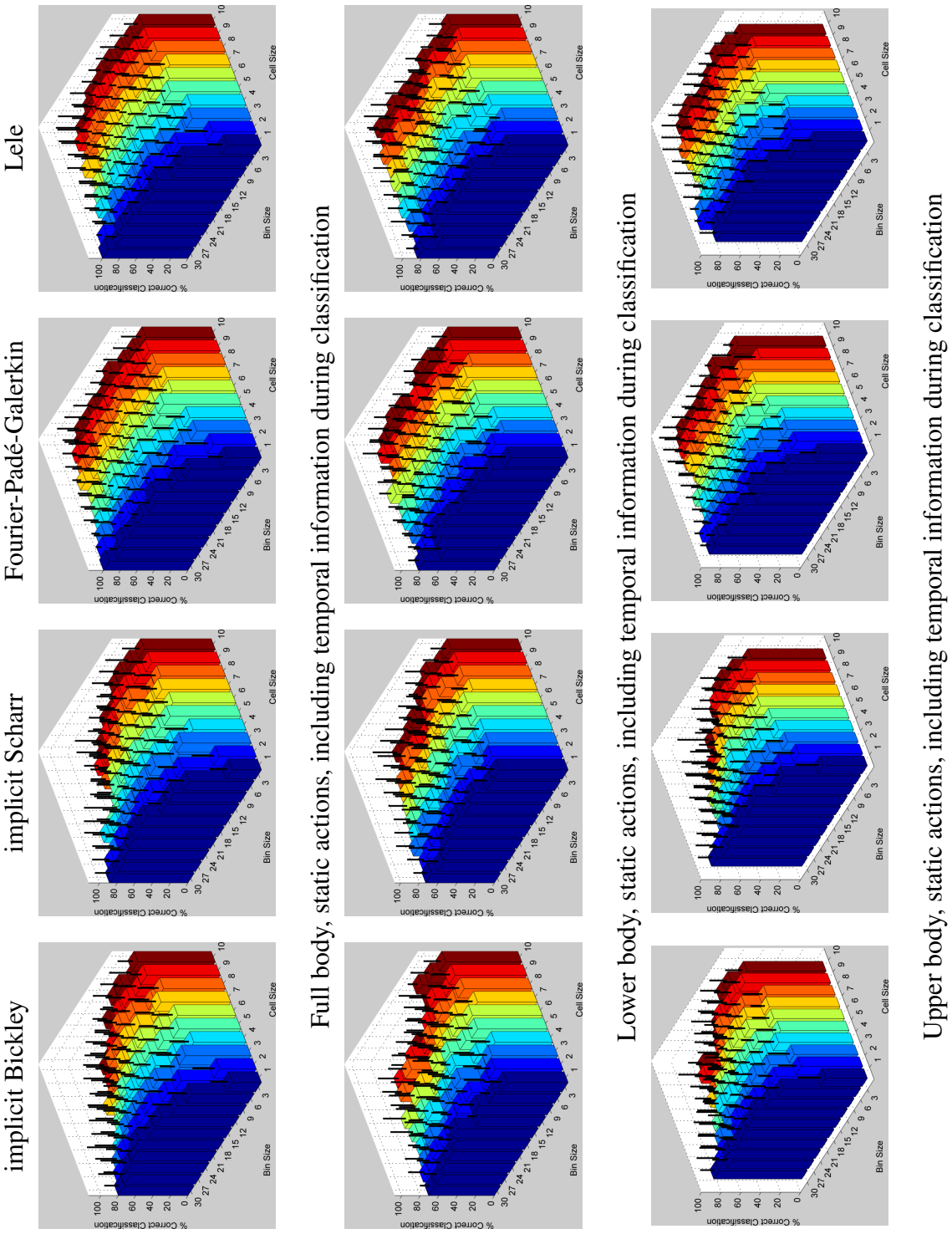


Figure 3.9 (Continued): Weizmann Action dataset normal action (5 classes) sequence performance with respect to HOG cell size and bin size, and body component contribution (full, lower and upper body) for each gradient scheme when incorporating temporal information by classifying static actions only; the error bars are shown in black

Classification Scheme (normal action sequences)	
With temporal information (%)	No temporal information (%)
93.5	89.3

Table 3.3: Weizmann Action dataset classification scheme performance averaged across body component contribution and gradient scheme

HOG cell size and bin size

The HOG cell size and bin size combinations are evaluated for each gradient scheme, body component contribution and classification scheme. Figure 3.7 shows the performance when no temporal information is considered during classification, while Figure 3.8 (dynamic action classes) and Figure 3.9 (static action classes) correspond to including temporal information during classification. Regardless of classification scheme and body component contribution, the following patterns exist with increasing cell size and bin size combinations. The performance of the explicit gradient schemes, Fourier-Padé-Galerkin scheme and Lele scheme increases prior to forming a plateau. The implicit Bickley and implicit Scharr scheme performance increases rapidly prior to decreasing.

The optimal HOG cell size and bin size for person detection is $c = 8, b = 9$. A Wilcoxon test ($p > .05$, two tailed test) indicates that action recognition results are significantly affected by the cell size and bin size. This occurs as the person detection HOG cell size and bin size combination is tuned to describe the width of a limb in a single image. However the GEI is a single compact 2D representation which cannot differentiate between unique limbs.

Classification scheme

Two distinct classification schemes are evaluated to determine how performance varies when incorporating temporal information during SVM classification. The first scheme (commonly used by action recognition approaches) classifies all (10 classes) actions. The second incorporates temporal information by dividing the actions into static (5 classes) and dynamic (5 classes) actions based on a global translation threshold. Therefore by incorporating temporal information, Table 3.3 shows performance increases by 4.7%. In addition, including temporal information during classification reduces action class num-

Weizmann Action dataset (normal action sequences %)

Gradient Scheme	Static Classes	Dynamic Classes	Average	Cell Size	Bin Size
centred (E)	97.8	100.0	98.9	5	27
Sobel (E)	95.6	100.0	97.8	4	18
explicit Bickley (E)	97.8	100.0	98.9	7	30
explicit Scharr (E)	97.8	100.0	98.9	7	30
implicit Bickley (I)	97.8	100.0	98.9	3	21
implicit Scharr (I)	97.8	100.0	98.9	3	24
Fourier-Padé-Galerkin (I)	100.0	100.0	100.0.0	4	12
Lele (I)	100.0	100.0	100.0.0	3	21

Table 3.4: Gradient scheme performance for the HOG cell size and bin size achieving the highest performance in normal action sequences; explicit and implicit gradient schemes are denoted by (E) and (I) respectively

one-handed wave GEI jump in place GEI

Figure 3.10: Misclassification occurs in static action classes due to the similarities between the one-handed wave GEI and jump in place GEI

bers by 50%; as a result, this reduces SVM classifiers by 77.8%. This positive result may benefit analogous applications with high class numbers e.g. gait recognition. This approach may be effective for robustness sequences as the candidate classes are reduced (this cannot be trialled with the Weizmann Action dataset as the robustness sequences are based only on the walking action).

General recommendations

The results in Table 3.1 and Table 3.3 indicate that full body GEIs and including temporal information during classification yields a high performance for action recognition. Therefore Table 3.4 shows the results for each gradient scheme given the highest performing combination of HOG cell size and bin size. The dynamic action classes achieve a high performance which indicates the combination of GEI and HOG can effectively distinguish between dynamic action classes. However misclassification occurs in the static

walk GEI at 36° jump GEI

Figure 3.11: Walking action GEIs which exceed 36° from the side view are commonly misclassified as jump GEIs

action classes between the visually similar one-handed wave GEI and jump in place GEI seen in Figure 3.10.

3.3.3.2 Robustness sequence evaluation

The robustness sequences are used to demonstrate how the performance varies when using the *a*) HOG cell size and bin size achieving the highest normal action sequence performance and *b*) HOG cell size and bin size achieving the highest robustness sequence performance. The robustness evaluation is somewhat limited as the sequences in the Weizmann Action dataset are based only on the walking action. However this is somewhat compensated by the quantity and quality of deformation and viewpoint sequences. The results seen in Table 3.5 are based on full body GEIs and including temporal information during classification.

HOG parameters achieving the highest normal sequence performance

Table 3.5 shows that these HOG parameters achieve a degree of robustness to deformations sequences (appearance-based and motion-based covariate factors), however the HOG parameters are sensitive to viewpoint sequences. A Wilcoxon test ($p > 0.05$, two-tailed test) indicates that *i*) compared to the centred scheme, the gradient schemes significantly affect the performance of deformation sequences, *ii*) compared to the centred scheme, the gradient schemes do not significantly affect the performance of viewpoint sequences and *iii*) the type (explicit/implicit) of gradient scheme does not significantly affect the performance of deformation and viewpoint sequences. Deformation sequence misclassification occurs with *i*) long term occlusion of dynamic features e.g. “walk with

Weizmann Action dataset - Deformation Sequences (%)

	using optimal HOG parameters for	
Gradient Scheme	normal action sequences	robustness sequences
centred (E)	70.0	90.0
Sobel (E)	90.0	100.0
explicit Bickley (E)	100.0	100.0
explicit Scharr (E)	100.0	100.0
implicit Bickley (I)	80.0	90.0
implicit Scharr (I)	80.0	90.0
Fourier-Padé-Galerkin (I)	70.0	90.0
Lele (I)	80.0	100.0

Weizmann Action dataset - Viewpoint Sequences (%)

	using optimal HOG parameters for	
Gradient Scheme	normal action sequences	robustness sequences
centred (E)	40.0	90.0
Sobel (E)	60.0	80.0
explicit Bickley (E)	50.0	100.0
explicit Scharr (E)	50.0	100.0
implicit Bickley (I)	30.0	90.0
implicit Scharr (I)	40.0	70.0
Fourier-Padé-Galerkin (I)	0.0	50.0
Lele (I)	30.0	70.0

Table 3.5: Robustness to deformation and viewpoint sequences using optimal HOG parameters for normal action sequences and robustness (deformation and viewpoint) sequences; these results are based on full body GEIs and using temporal information during classification

Weizmann Action dataset (%)

Gradient Scheme	optimal HOG parameter for normal action sequences			optimal HOG parameter for robustness sequences		
	Dynamic Classes	Bin Size	Cell Size	Dynamic Classes	Bin Size	Cell Size
centred (E)	100.0	5	27	97.8	10	27
Sobel (E)	100.0	4	18	88.9	8	12
explicit Bickley (E)	100.0	7	30	97.8	10	15
explicit Scharr (E)	100.0	7	30	97.8	10	15
implicit Bickley (I)	100.0	3	21	80.0	9	9
implicit Scharr (I)	100.0	3	24	84.4	9	6
Fourier-Padé-Galerkin (I)	100.0	4	12	97.8	3	9
Lele (I)	100.0	3	21	93.3	6	27

Table 3.6: HOG cell size and bin size requirements to achieve the highest normal action sequence performance and highest robustness sequence performance

jump GEI skip GEI

Figure 3.12: Dynamic action class misclassification occurs between the jump GEI and skip GEI

a dog" and "occluded feet" and *ii*) unnatural walking patterns e.g. "moonwalk" and "limp walk". Viewpoint sequence performance rapidly decreases when the view exceeds 36° from the side view. As seen in Figure 3.11, these views are commonly misclassified as the jump action.

HOG parameters achieving the highest robustness sequence performance

Table 3.5 shows that HOG parameters achieving the highest normal sequence performance are not effective for robustness sequence performance. By selecting HOG parameters achieving the highest performance during deformation and viewpoint sequences, the robustness can significantly increase. A Wilcoxon test ($p > 0.05$, two-tailed test) indicates that *a*) compared to the centred scheme, gradient schemes do not significantly affect the performance of deformation and viewpoint sequences and *b*) the type (explicit/im-

Weizmann Action dataset (%)			
Explicit schemes (%)		Implicit schemes (%)	
centred	-2.2	implicit Bickley	-20.0
Sobel	-11.1	implicit Scharr	-15.6
explicit Bickley	-2.2	Fourier-Padé-Galerkin	-2.2
explicit Scharr	-2.2	Lele	-6.7

Table 3.7: Normal action sequence performance decreases when using HOG parameters achieving the highest performance in robustness sequences

jump GEI skip GEI run GEI

Figure 3.13: The jump, skip and run actions are commonly misclassified in action recognition

plicit) of gradient scheme does not significantly affect the performance of deformation and viewpoint sequences. As seen in Table 3.6, gradient schemes require the bin size to be increased and the cell size to be decreased. This indicates that robustness is achieved by describing a smaller GEI area with finer binning (decreased cell size and increased bin size). The exception to this is the *a*) Fourier-Padé-Galerkin scheme which requires decreased cell size and bin size and *b*) Lele scheme which requires increased cell size and bin size. These differences may be cause by the high gradient orientation and magnitude accuracy compared to the other gradient schemes seen in Figure 3.3. The trade-off for robustness can be seen in Table 3.7 where these HOG parameter values cause a decrease in normal action sequence performance. The misclassification in dynamic action classes occurs between jump GEIs and skip GEIs seen in Figure 3.12.

The robustness sequence results mirror the results achieved in normal action sequences whereby *i*) action recognition, like person detection, requires a higher gradient orientation accuracy compared to gradient magnitude accuracy and *ii*) the optimal person detection HOG parameters are not effective for action recognition.

Weizmann Action dataset (normal dynamic class action sequences %)

Action Representation	jump	run	gallop sideways	skip	walk	average
Space-time shapes Gorelick et al. (2007a)	89.2	98.0	**100.0**	97.1	**100.0**	96.9
HOOF Chaudhry et al. (2009b)	89.0	89.0	**100.0**	**100.0**	**100.0**	95.6
GEI and HOG (proposed)	**100.0**	**100.0**	**100.0**	88.9	**100.0**	**97.8**

Table 3.8: Comparison of the GEI and HOG combination to state-of-the-art results in the Weizmann action dataset; the results are based on the normal dynamic class action sequences

3.3.4 Comparison to State-of-the-Art

A recent benchmark [Liu et al. (2011)] in the Weizmann Action dataset normal action sequences shows 100% performance [Fathi and Mori (2008),Tran and Sorokin (2008b),Wang and Mori (2009),Yeffet and Wolf (2009)]. However the benchmark lacks robustness sequence performance which is of greater importance. Comparing the result in this chapter is not strictly fair due to the alternative classification scheme which divides action classes prior to classification. Therefore comparison in Table 3.8 is based on the normal dynamic action class only. The proposed GEI and HOG combination uses *i*) full body GEIs, *ii*) including temporal information during classification and *iii*) the explicit Bickley/Scharr gradient schemes with HOG parameters achieving the highest robustness sequence performance. Gorelick et al. (2007a) robustness sequence performance is matched at 100% (Chaudhry et al. (2009b) do not use these sequences). The proposed GEI and HOG combination achieves a 0.9% performance increase compared to Gorelick et al. (2007a), and a 2.3% performance increase against Chaudhry et al. (2009b). Figure 3.13 shows the common confusion between jump, skip and run actions.

3.3.5 Conclusion

The combination of GEI and HOG is successful for action recognition due to three primary reasons.

1. Compared to the centred scheme, a Wilcoxon test ($p > 0.05$, two-tailed test) indicates that gradient schemes do not significantly affect normal action and robustness sequence performance. In addition, a Wilcoxon test ($p > 0.05$, two-tailed test) indicates that the type (explicit/implicit) of gradient scheme does not significantly affect normal and and robustness sequence performance. The optimal cell size and bin size for person detection are used describe the width of a limb. Compared to person detection which requires $c = 8, b = 9$, a Wilcoxon test ($p > .05$, two tailed test) indicates that action recognition results are significantly affected by the cell size and bin size. The optimal person detection cell size and bin size are not effective for action recognition as the GEI cannot visually distinguish unique limbs.

2. Action GEIs contain a varied distribution of static features and dynamic features (compared to gait GEIs which show a regular distribution due to only considering the walking action). Therefore static features and dynamic features are discriminative for action recognition meaning a high performance is achieved when using full body GEIs.

3. Temporal information somewhat lacks in the GEI due to condensing a silhouette sequence into a single compact 2D representation. This is alleviated by incorporating temporal information in the classification stage. Therefore action classes are divided into static action classes and dynamic action classes based on a threshold of global translation. This technique halves the candidate action classes and reduces the required SVM classifiers by 77.8%.

Hypothesis Revised

At the beginning of this chapter, the following hypothesis was made:

> "This chapter argues that to ensure robustness, HOG parameters (gradient scheme, cell size and bin size) must undergo re-evaluation when applied to applications other than person detection."

Therefore during the course of this chapter, the hypothesis has been verified.

Future Directions

Dividing the action classes in action recognition achieves a higher performance. Therefore gender recognition or age recognition could be used to increase gait recognition performance by dividing the person classes. Note that these topics are separate research areas in their own right and are not implemented in this thesis.

3.4 Application: Gait Recognition

The combination of GEI and HOG exists for gait recognition [Sun et al. (2010),Liu et al. (2012)]. However these results indicate a high performance can only be achieved when using feature fusion or a small validation dataset. Sun et al. (2010) perform validation in the CMU MoBo dataset which contains 25 persons, however this chapter uses the CASIA B and TUM GAID datasets which contain 124 and 155 test persons respectively.

The analysis of body component contribution is not required for gait recognition as this has been established. Early research by Veeraraghavan et al. (2004) and Veres et al. (2004) suggest that static features (upper body) are more discriminative for gait recognition. However recent research by Bashir et al. (2008a) and Martin-Félez et al. (2010) show that dynamic features (lower body) are discriminative and robust to covariate factors. Section 3.3.3.1 shows that action recognition results are increased by incorporating temporal information during classification. For gait recognition, this could be implemented by performing gender recognition or age recognition. Note that these research topics are outwith this thesis and are therefore suggestions for future work.

3.4.1 Experimental Procedure: Gait Recognition

3.4.1.1 GEI construction

The GEI requires space-normalisation prior to (3.1). This is achieved by horizontally aligning the centroid of the top 10% of the silhouette (head) as a reference point. This process is similar to Hofmann et al. (2011) and the same as Section 3.3.2.1.

3.4.1.2 Dataset

The combination of GEI and HOG is validated in the CASIA B and TUM GAID datasets which are explained in Chapter 2.4.

3.4.1.3 HOG parameters

Eight gradient schemes are evaluated: *1*) centred, *2*) Sobel, *3*) explicit Bickley, *4*) explicit Scharr, *5*) implicit Bickley, *6*) implicit Scharr, *7*) Fourier-Padé-Galerikin and *8*) Lele. One hundred combinations of cell size and bin size are evaluated (note that larger cell sizes and bin sizes yield high dimensionality feature vectors). Cell sizes c range from $\{c = 1$ in steps of 1 to $c = 10\}$ and bin sizes b range from $\{b = 3$ in steps of 3 to $b = 30\}$.

3.4.1.4 Classification

As gait recognition is a multi-class problem, this requires a more complex classification approach (compared to person detection which is a simple two-class problem of person versus not a person). Multi-class SVM can be used with one-versus-all (OVA) or one-versus-one (OVO) binary classification. OVA is heuristic to some degree [Scholkopf and Smola (2001)] as a winner-takes-all approach. However OVO requires more classifiers than OVA which increases the computational demand. Initial experiments demonstrate that OVO significantly outperforms OVA. This is common [Hsu and Lin (2002)] and therefore subsequent results are based on OVO. The OVO SVM results are compared against the ground truth in a confusion matrix. The average of the diagonal in the confusion matrix yields the correct classification.

3.4.2 Results and Discussion

The CASIA B and TUM GAID dataset results for each gradient scheme is shown in Table 3.9 for the cell size and bin size achieving the highest performance across all covariate factors (this is standard for gait recognition). The results for each HOG cell size and bin size combination is seen in Figure 3.14 and Figure 3.15 for the CASIA B and TUM GAID datasets respectively. Given these results, the performance is discussed based on the *i*) gradient scheme and *ii*) HOG cell size and bin size.

CASIA B dataset (124 test persons)

Gradient Scheme (%)	nm	bg	cl	avg	cell size	bin size
centred (E)	96.4	41.5	25.8	**54.6**	24	7
Sobel (E)	**97.2**	33.1	24.2	51.5	18	5
exp. Bickley (E)	88.3	**45.6**	29.8	**54.6**	18	7
exp. Scharr (E)	96.8	35.9	27.0	53.2	21	8
imp. Bickley (I)	88.3	**45.6**	29.8	**54.6**	18	7
imp. Scharr (I)	94.0	27.8	13.7	45.2	18	4
Fourier-Padé-Galerkin (I)	89.5	29.8	**30.6**	50.0	3	10
Lele (I)	87.1	24.6	27.4	46.4	3	8

TUM GAID dataset (155 test persons)

Gradient Scheme (%)	N	B	S	TN	TB	TS	avg	cell size	bin size
centred (E)	95.5	14.2	88.4	**40.6**	3.1	34.4	**62.3**	24	10
Sobel (E)	89.4	13.9	69.4	25.0	3.1	31.3	54.0	30	3
exp. Bickley (E)	72.3	11.9	57.1	15.6	9.4	21.9	44.2	30	8
exp. Scharr (E)	**96.1**	10.6	**89.7**	37.5	3.1	**46.9**	62.1	30	10
imp. Bickley (I)	72.3	11.9	57.1	15.6	9.4	21.9	44.2	30	8
imp. Scharr (I)	89.4	14.5	65.8	25.0	6.3	40.6	53.5	27	3
Fourier-Padé -Galerkin (I)	88.1	**16.5**	74.5	15.6	**18.8**	21.9	55.8	3	8
Lele (I)	85.2	11.9	71.3	18.8	15.6	12.5	52.3	3	8

Table 3.9: Gradient scheme performance in the CASIA B and TUM GAID datasets for the HOG cell size and bin size achieving the highest average performance across all covariate factors; explicit and implicit gradient schemes are denoted by (E) and (I) respectively. CASIA B dataset sequences: normal (nm), carrying a bag (bg), clothing (cl). TUM GAID dataset sequences: normal (N), carrying a bag (B), shoes (S), time and normal (TN), time and carrying a bag (TB), time and shoes (TS)

3.4.2.1 Covariate factor performance trends

Table 3.9 shows that covariate factor free GEIs (CASIA B: nm, TUM GAID: N) achieve a high performance due to their visual similarities with training GEIs. However covariate factor GEI (CASIA B: bg, cl, TUM GAID: B, S, TN, TB, TS) performance is poorer due to the unique ways in which covariate factors affect the natural appearance and motion of gait. The time-based covariate factor GEIs in the TUM GAID dataset show a poor performance due to the complex coupled covariate factors, i.e. clothing as well as the named covariate factor. Note that the variation in CASIA B and TUM GAID bag sequence performance is linked to the varying bag types. The CASIA B dataset uses varying bag types (handbags, rucksacks, satchels etc. which vary in location on the body) while the TUM GAID dataset uses a consistent bag (rucksack).

3.4.2.2 Gradient scheme

The CASIA B dataset gradient scheme results are shown in Table 3.9. A Wilcoxon test ($p > 0.05$, two-tailed test) indicates that compared to the centred scheme, gradient schemes do not significantly affect the performance in each sequence type (normal, carrying a bag, clothing). In addition, a Wilcoxon test ($p > 0.05$, two-tailed test) indicates that gait recognition performance is not significantly affected by the type (explicit/implicit) of gradient scheme.

Table 3.9 shows the TUM GAID dataset gradient scheme results. A Wilcoxon test ($p > 0.05$, two-tailed test) indicates that compared to the centred scheme, gradient schemes significantly affect the performance in normal, shoes, time and normal, and time and bag sequences; the remaining sequences (bag, time and shoes) are not significantly affected by the gradient scheme. Across the sequences, a Wilcoxon test ($p > 0.05$, two-tailed test) indicates that gait recognition performance is not significantly affected by the type (explicit/implicit) of gradient scheme. For both datasets, the higher performing explicit schemes indicate that gait recognition requires higher gradient orientation accuracy compared to gradient magnitude accuracy. This observation is consistent with person detection and action recognition results.

The CASIA B and TUM GAID dataset results show there is no single gradient scheme

which is effective for every sequence type. This is due to the unique way in which covariate factors affect the natural appearance and motion of gait. While the CASIA B and TUM GAID datasets use normal (covariate factor free) and bag sequences, the highest performing gradient schemes do not match. This may be attributed to the *i*) image size (standard dataset image sizes CASIA B: 240×240, TUM GAID: 128×178), *ii*) silhouette quality, given the TUM GAID dataset extracts higher quality silhouettes based on depth images compared to the CASIA B which uses background subtraction and *iii*) the different bags used in the datasets.

3.4.2.3 HOG cell size and bin size

The HOG cell size and bin size combinations are evaluated for each gradient scheme. CASIA B and TUM GAID dataset results are shown in Figure 3.14 and Figure 3.15 respectively.

Regardless of dataset, the following performance patterns exist with increasing cell size and bin size. The explicit (and implicit Bickley) scheme performance increases prior to forming a plateau and subsequently deceasing. The implicit Scharr scheme mimics this pattern in a faster manner. Conversely, the Fourier-Padé-Galerkin and Lele schemes achieve the highest performance during large cell size and bin size and performance decreases thereafter.

There is no single cell size and bin size which yields equally high performance across sequence type. This is due to the unique way in which covariate factors affect the natural appearance and motion of gait. The person detection HOG cell size and bin size combination is chosen to describe the width of a limb. However the GEI is a single compact 2D gait representation which cannot differentiate between unique limbs. Compared to person detection which requires $c = 8, b = 9$, a Wilcoxon test ($p > .05$, two tailed test) indicates that the *a*) CASIA B dataset results are significantly affected by the bin size but not cell size and *b*) TUM GAID dataset results are not significantly affected by the cell size and bin size.

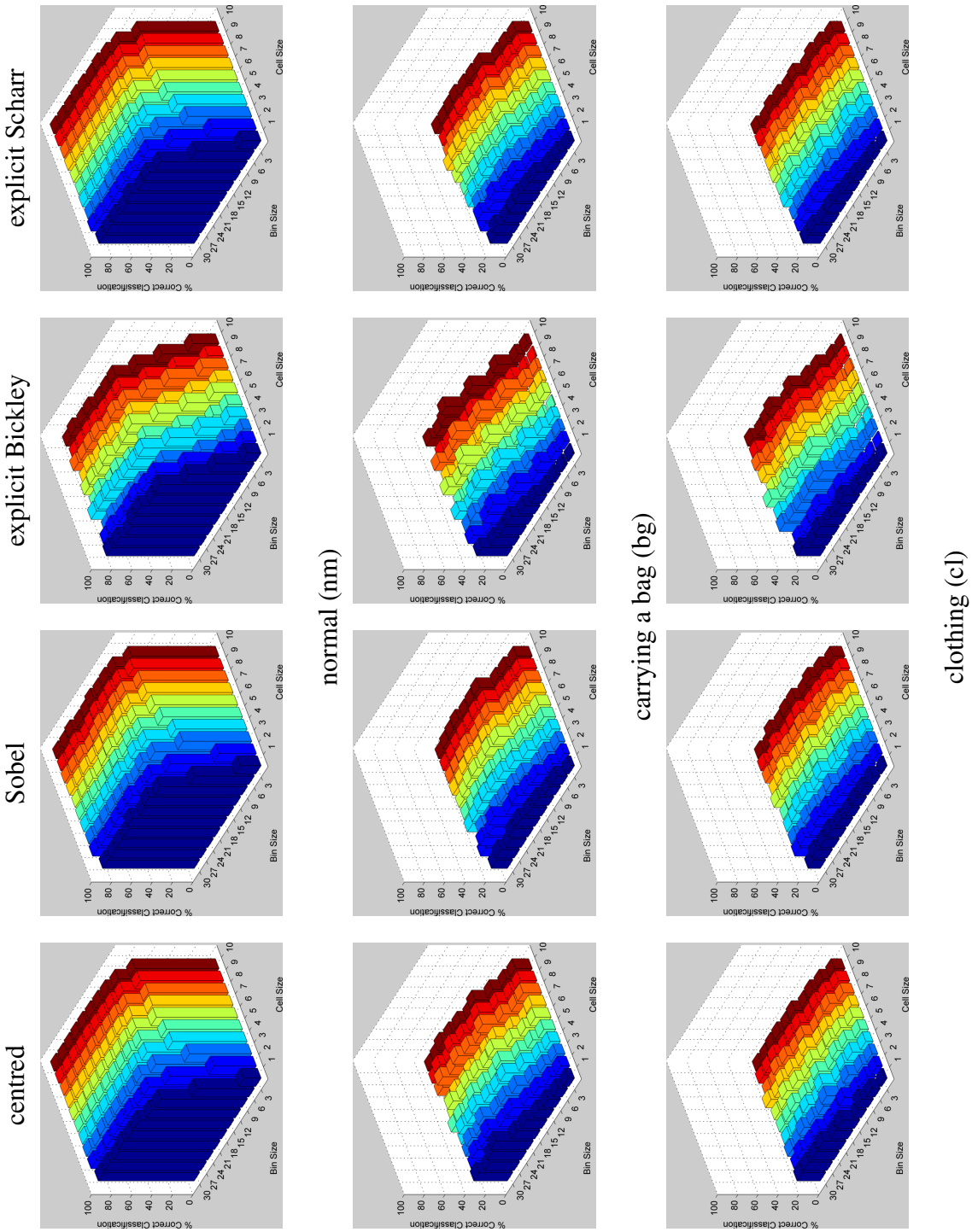


Figure 3.14: CASIA B dataset (124 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

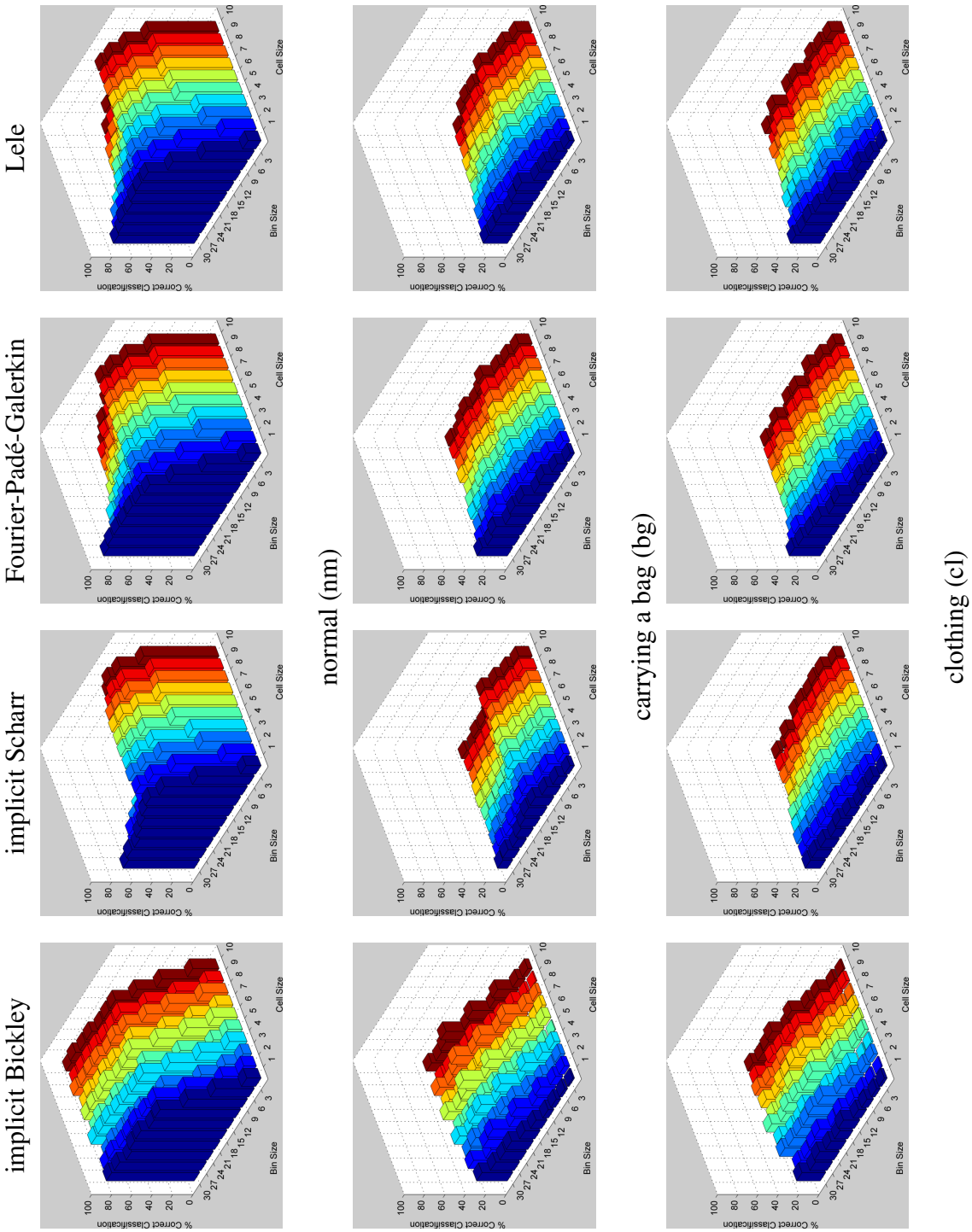


Figure 3.14 (Continued): CASIA B dataset (124 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

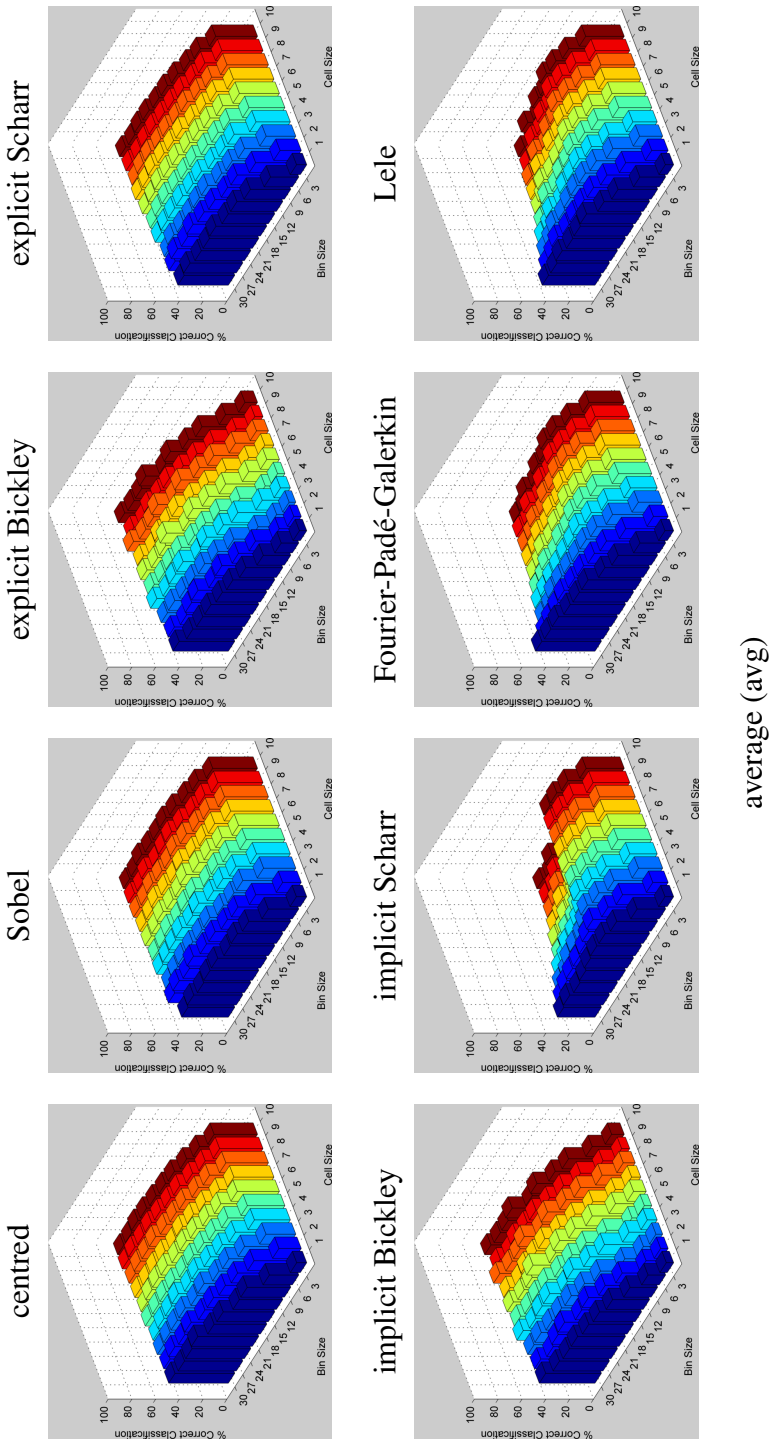


Figure 3.14 (Continued): CASIA B dataset (124 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

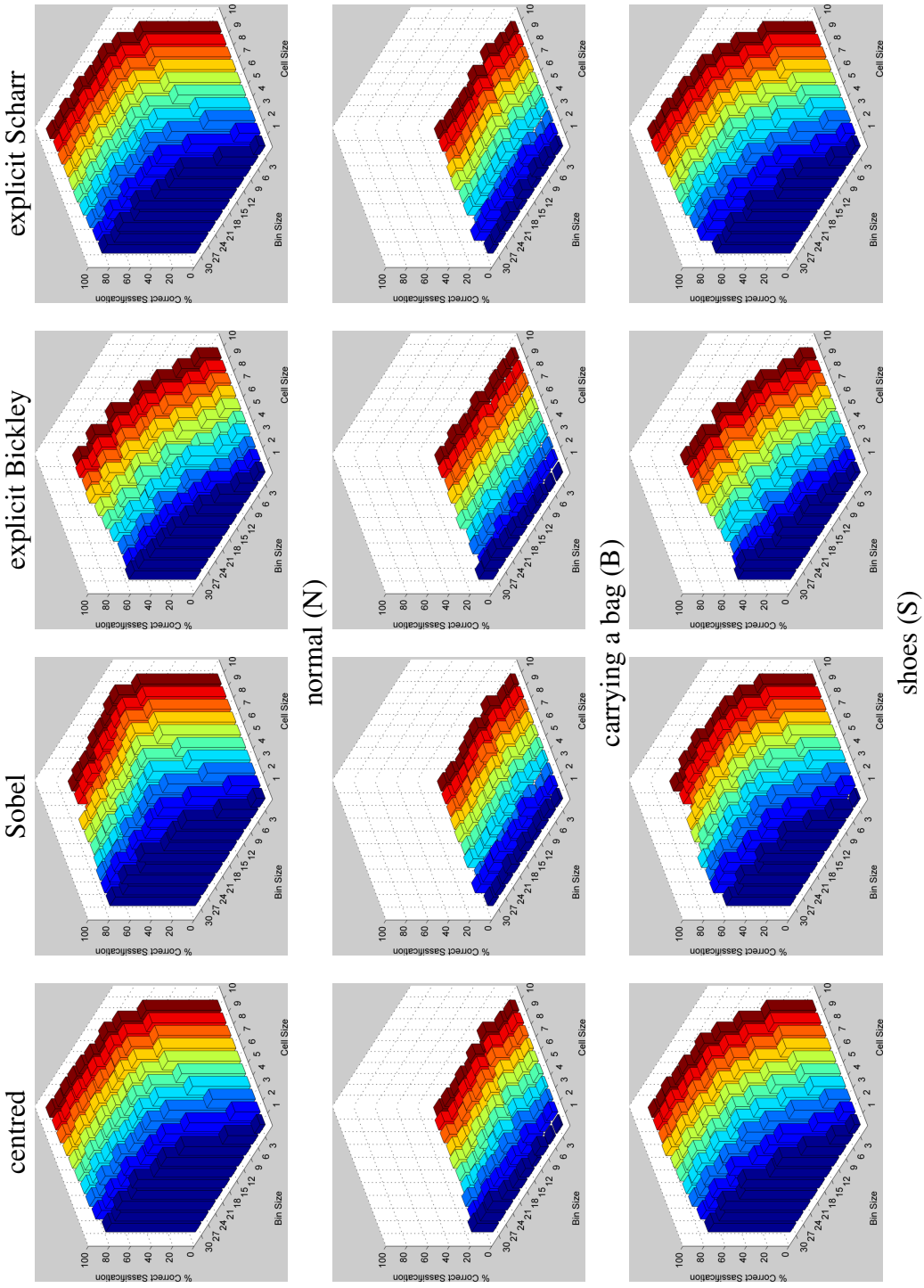


Figure 3.15: TUM GAID dataset (155 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

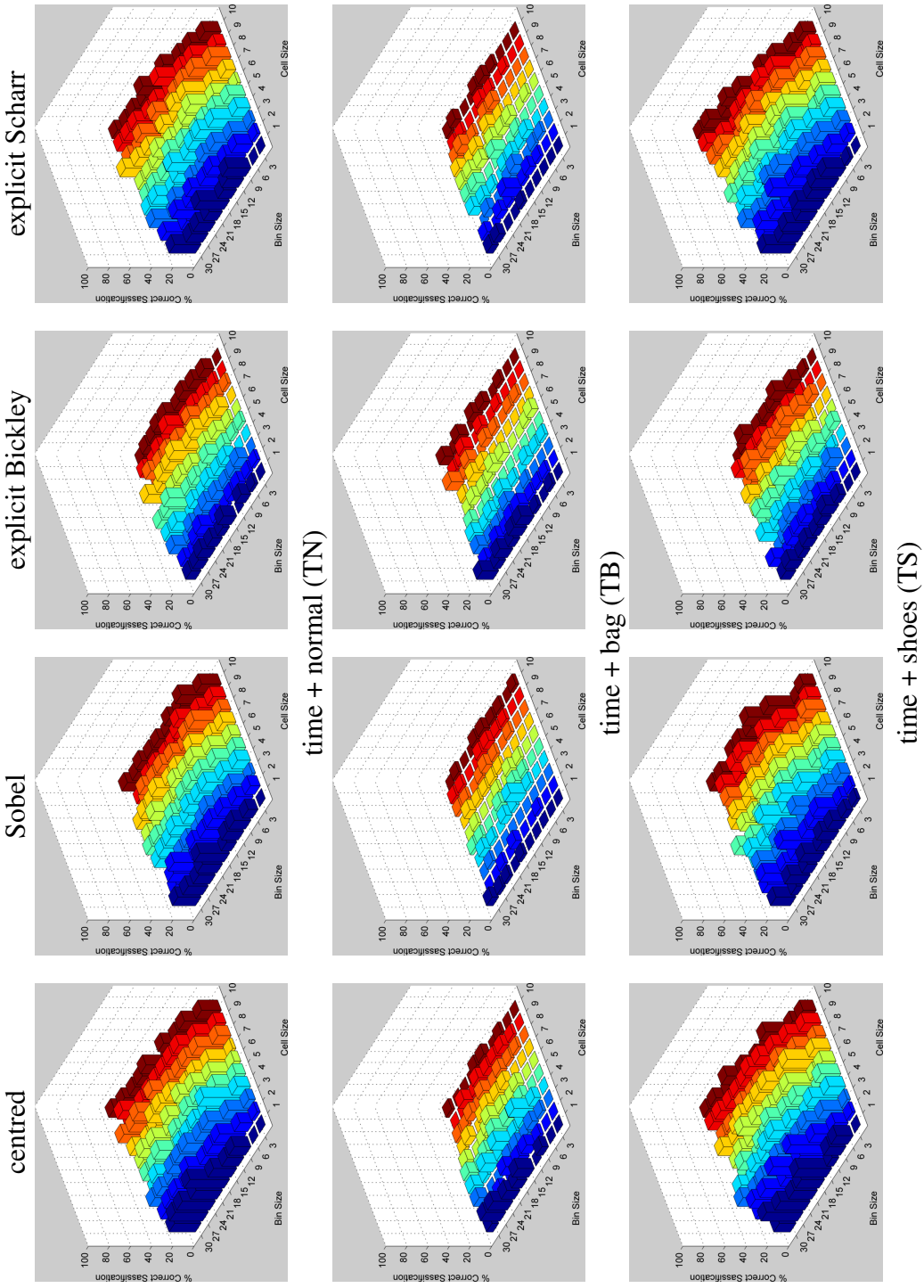


Figure 3.15 (Continued): TUM GAID dataset (155 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

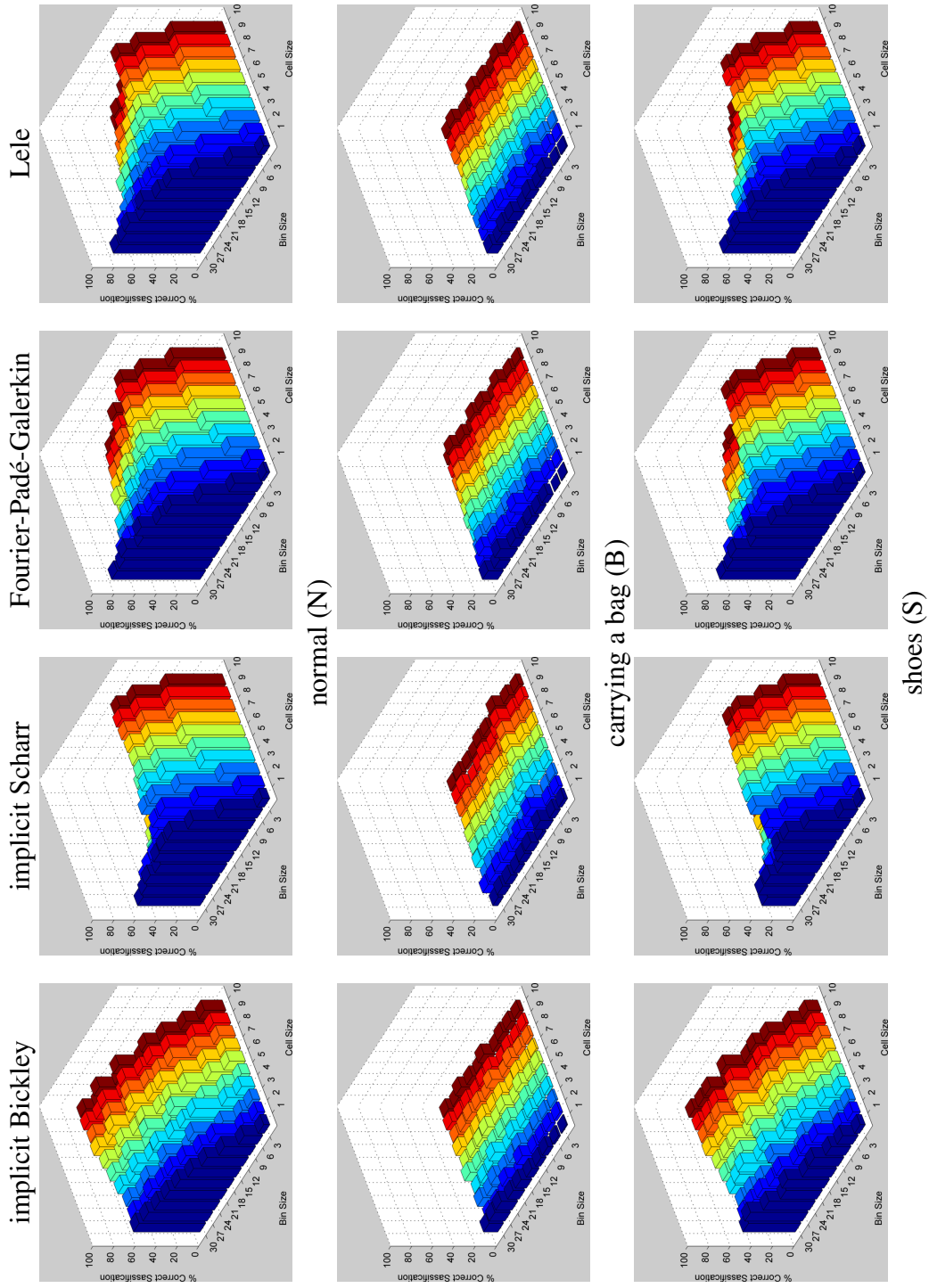


Figure 3.15 (Continued): TUM GAID dataset (155 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

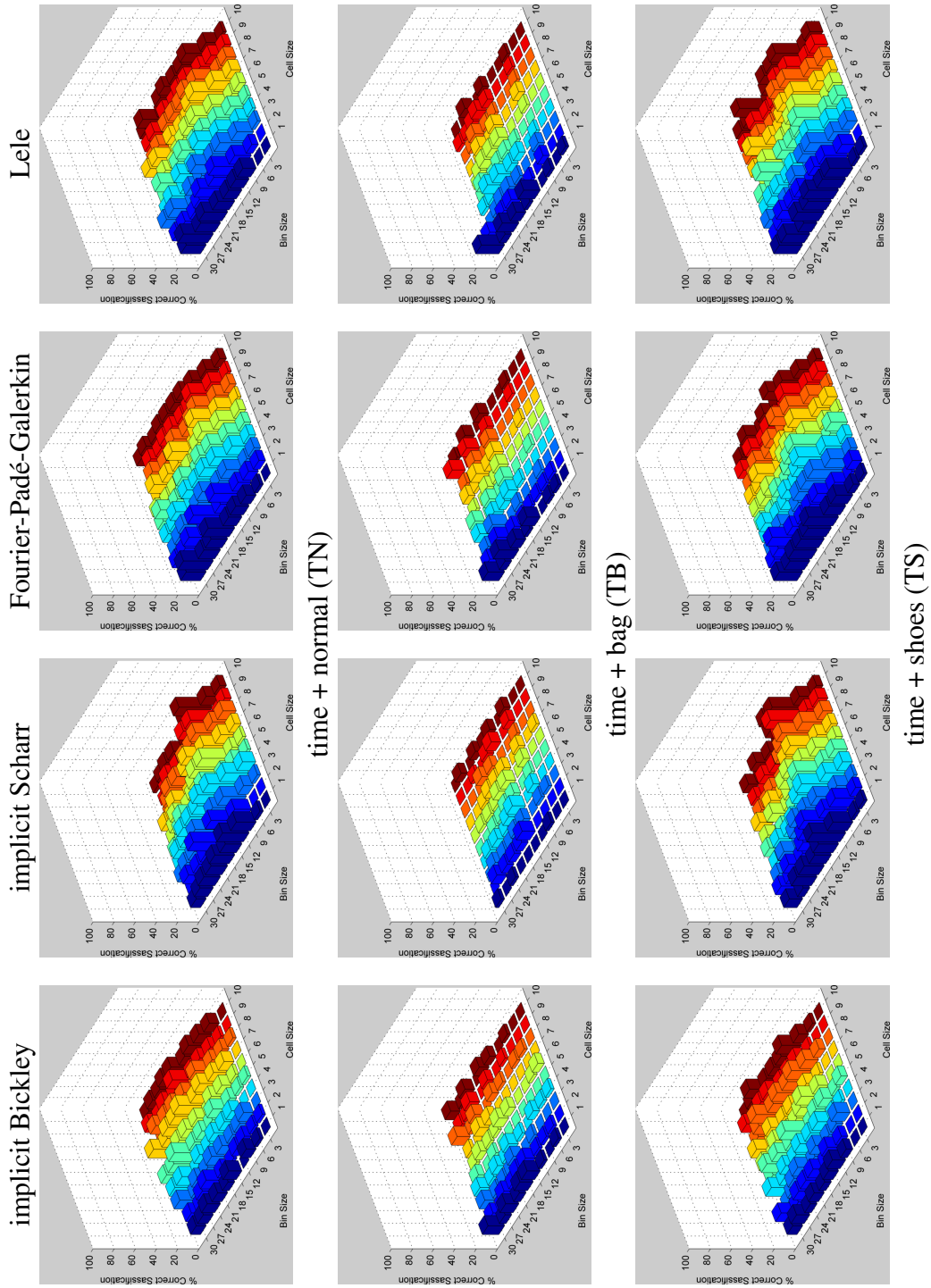


Figure 3.15 (Continued): TUM GAID dataset (155 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

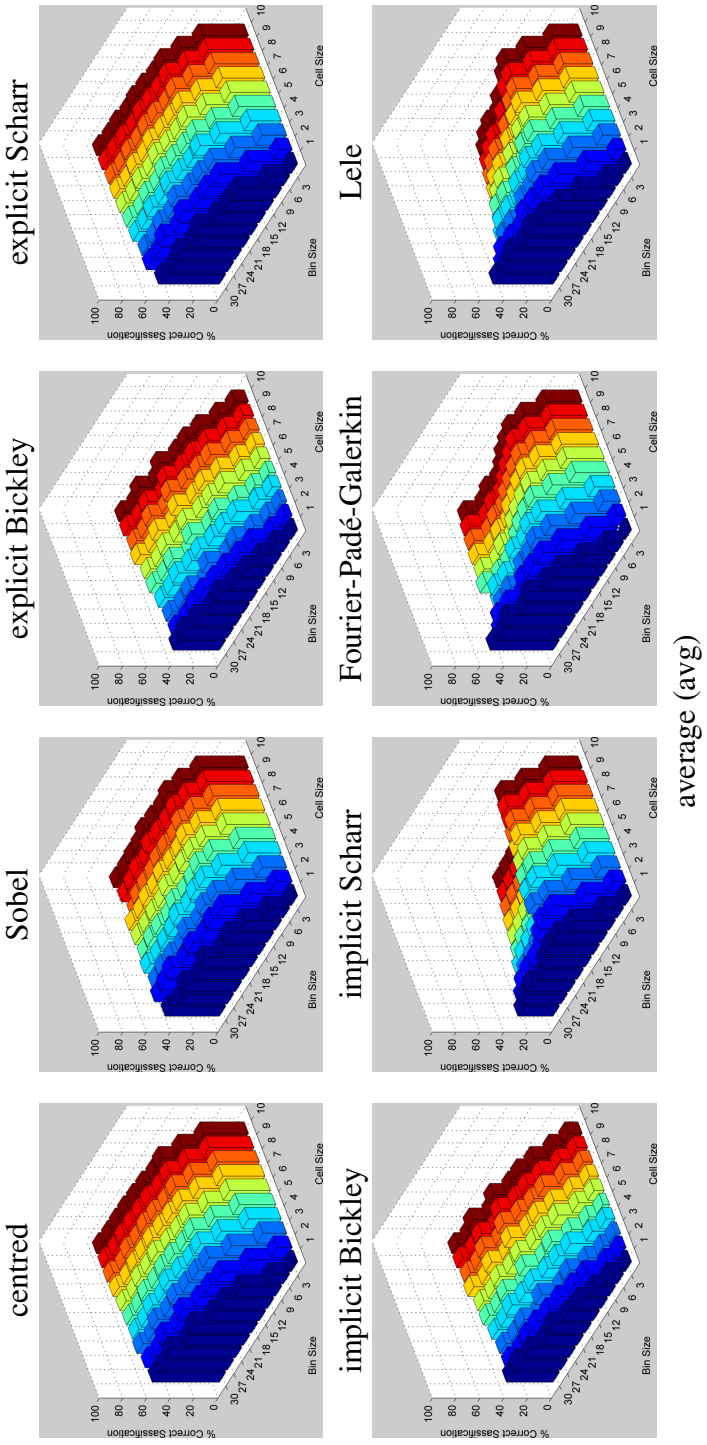


Figure 3.15 (Continued): TUM GAID dataset (155 test persons) results with respect to HOG cell size and bin size for each gradient scheme and sequence

CASIA B dataset (124 test persons)

Gait Representation (%)	nm	bg	cl	avg
CGI Wang et al. (2012)	88.1	43.7	43.0	58.2
GEI Han and Bhanu (2006)	100.0	53.2	22.2	58.5
GEnI Bashir et al. (2010)	**100.0**	78.3	44.0	74.1
MII + MDIs Bashir et al. (2009b)	97.5	83.6	48.8	76.6
SEI + GSP Huang and Boulgouris (2012)	99.0	64.0	72.0	78.3
P_{RW}GEI Yogarajah et al. (2011)	98.4	**93.1**	44.4	78.6
Body segmentation Li et al. (2010)	99.2	80.6	75.8	85.2
AEI Zhang et al. (2010)	98.4	91.9	72.2	87.5
SGEI + GEI Li and Chen (2013)	98.2	80.7	**83.9**	87.6
M_G Bashir et al. (2008a)	**100.0**	91.0	80.6	**90.5**
GEI + HOG (proposed) (centred scheme)	96.4	41.5	25.8	54.6
GEI + HOG (proposed) (imp/exp Bickley scheme)	88.3	45.6	29.8	54.6

Table 3.10: Comparison of the GEI and HOG results to current state-of-the-art results in the CASIA B dataset for normal (nm), carrying a bag (bg) and clothing (cl) sequences

TUM GAID dataset (155 test persons)

Gait Representation (%)	N	B	S	TN	TB	TS	average
depth GEI Han and Bhanu (2006)	**99.7**	17.4	**96.5**	37.5	0.0	43.8	67.1
GEV Hofmann et al. (2013)	94.2	13.9	87.7	41.0	0.0	31.0	61.4
DGHEI Hofmann et al. (2013)	99.0	**40.3**	96.1	**50.0**	0.0	**44.0**	**74.1**
GEI + HOG (proposed) (centred scheme)	95.5	14.2	88.4	40.6	**3.1**	34.4	62.3

Table 3.11: Comparison of the GEI and HOG results to current state-of-the-art results in the TUM GAID dataset for normal (N), carrying a bag (B), shoes (S), time and normal (TN), time and bag (TB), time and shoes (TS) sequences

3.4.3 Comparison to State-of-the-Art

The proposed approach of HOG describing the GEI is compared to state-of-the-art in Table 3.10 and Table 3.11 for the CASIA B and TUM GAID datasets respectively. Notice that the TUM GAID dataset achieves a higher performance compared to the CASIA B dataset. This is attributed to the difference in silhouette quality between the datasets. The TUM GAID dataset uses the Microsoft Kinect to extract relatively clean and intact silhouettes. However the CASIA B dataset uses background subtraction which yields imperfect silhouettes containing noise which causes missing heads or limbs.

Despite evaluating a variety of gradient schemes ranging in gradient orientation and magnitude accuracy, Table 3.10 and Table 3.11 show that the GEI and HOG combination does not exceed state-of-the-art results. While the performance during normal (covariate factor free) sequences is satisfactory, GEI and HOG cannot generalise over the covariate factor sequences. This is attributed to HOG encoding the covariate factors in the GEI. Therefore these results suggest that the combination of GEI and HOG is only effective during a feature fusion approach [Sun et al. (2010),Liu et al. (2012)], validation in small datasets [Sun et al. (2010)] or combining GEI and HOG in a different manner [Hofmann et al. (2013)]. Hofmann et al. (2013) present the Depth Gradient Histogram Energy Image (DGHEI) which is the average of HOG vectors applied to each depth image in the depth image sequence.

3.4.4 Conclusion

Describing the GEI with HOG for gait recognition yields unsatisfactory results due to two primary reasons.

1. The combination of GEI and HOG achieves a satisfactory performance in normal (covariate factor free) sequences. However the combination is ineffective for covariate factor sequences. This is due to HOG encoding the covariate factor appearance in the GEI.

2. The combination of GEI and HOG does not scale with increasing person numbers in a dataset.

3. A Wilcoxon test ($p > .05$, two tailed test) indicates that gait recognition results can be significantly affected by the gradient scheme (depending on the covariate factor type) and cell size. The gradient scheme type (explicit/implicit) and bin size do not significantly affect gait recognition results.

Hypothesis Revised

At the beginning of this chapter, the following hypothesis was made:

> "This chapter argues that to ensure robustness, HOG parameters (gradient scheme, cell size and bin size) must undergo re-evaluation when applied to applications other than person detection."

Therefore during the course of this chapter, the hypothesis has been verified.

Future Directions

- Action recognition performance is increased by reducing the action classes during classification. Therefore the same approach may benefit gait recognition. This could be achieved by age recognition or gender recognition.

The GEI and HOG combination is ineffective for covariate factor sequences as HOG encodes the appearance of covariate factors in the GEI. Therefore Chapter 4 is devoted to mitigating the effects of covariate factors in single compact 2D gait representations.

Chapter 4

Variance-based Fuzzy Skeletal Features

This chapter is devoted to exploiting the novel combination of skeleton representations and single compact 2D gait representations. Skeletons are infrequently used for gait recognition due to their sensitivity to boundary noise caused by imperfect extraction. This sensitivity can be alleviated by using a smooth distance function to absorb boundary noise. Therefore three smooth distance functions are trialled for their varying accuracy properties; this is essential to determine the properties required for robust gait recognition. Fuzzy skeletons are extracted from the smooth distance function by computing the gradient. The term fuzzy refers to the thickness of the skeleton compared to the true skeleton. The fuzzy skeleton sequence is condensed into the single compact 2D Skeleton Variance Image (SVIM) gait representation by computing the pixel-wise variance. This process expresses motion features which are less sensitive to covariate factors. The SVIM is a powerful gait descriptor which yields a 9.9% increase over current state-of-the-art results when validated in the TUM GAID dataset.

Hypothesis

This chapter argues that by exploiting the Poisson equation to construct a smooth distance function, fuzzy skeletons can be extracted and formed into a single compact 2D gait representation to yield a discriminative gait descriptor.

Publications

The results of this chapter have been published in the Journal of Mathematical Imaging and Vision [Whytock et al. (2014)].

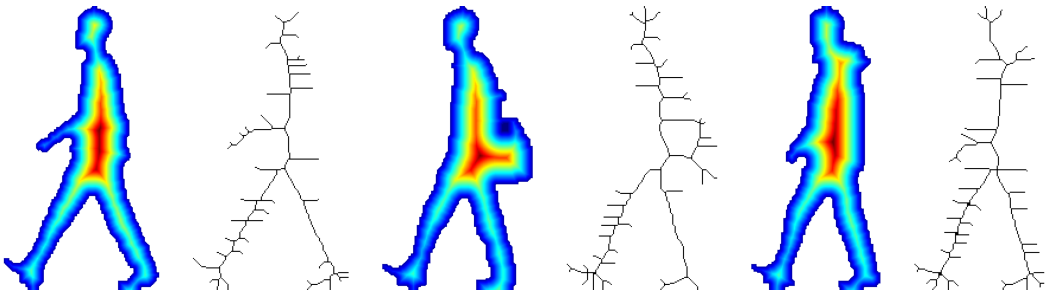
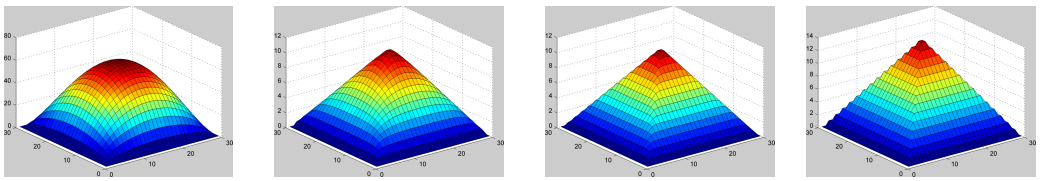


Figure 4.1: The distance function (left in pair) generated by the Euclidean metric demonstrates the retention of boundary noise across cool and hot colours. The corresponding skeleton extracted by the medial axis transform (right in pair) shows additional unwanted spurs

Since the pioneering work of Blum (1967), skeleton representations have gained popularity within computer vision, pattern recognition, image processing, shape matching and computer graphics applications. While the process of representing a domain (enclosed shape) by a skeleton is a well established technique, skeletons are seldom used for gait recognition due to the *a*) boundary noise sensitivity caused by imperfect silhouette extraction and *b*) the natural self-occluding nature of gait. Instead gait recognition favours a stick figure which has the advantages of *a*) reduced sensitivity to boundary noise and *b*) the ability to distinguish between unique legs during periods of self-occlusion. A notable example of stick figures in gait recognition is by Yoo and Nixon (2011) where anthropometrics define the locations of key human body joints (neck, shoulder, waist, pelvis, knee and ankle) which are connected to form a stick figure.

A skeleton can be extracted from the distance function or medial axis transform. The distance function is a scalar (vector) field which approximates the minimum distance (and direction) from a point inside a domain to the boundary. The medial axis transform is closely related to the distance function from the domain boundary (the medial axis transform can be defined as the set of singularities of the distance function). The skeletons in Figure 4.1 demonstrate numerous additional spurs caused by extraneous boundary noise resulting from imperfect silhouette extraction. A smooth distance function can be used to absorb boundary noise and yield a skeleton [Aubert and Aujol (2012),Direkoglu et al. (2012),Gorelick et al. (2006),Tari et al. (1997)] which is free from anomalous spurs.

To this end, this chapter is devoted to *1*) extracting newly defined fuzzy skeletons (the term fuzzy is applied due to the skeleton thickness compared to the true skeleton)



Poisson normalised Poisson screened Poisson true

Figure 4.2: Smooth distance functions generated from the Poisson (4.1), normalised Poisson (4.1), (4.2) and screened Poisson (4.4), (4.6) equations are compared against the true distance function for a simple square domain. Notice how the approximation accuracy progressively increases from left to right

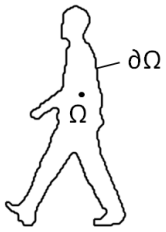


Figure 4.3: Geometry of the distance function

derived from the smooth distance function which alleviates boundary noise sensitivity and *2*) exploiting the gap in knowledge relating to the combination of skeleton and single compact 2D gait representations.

4.1 Smooth Distance Function

This chapter focusses on smooth distance functions derived from partial differential equations [Tari et al. (1997)], specifically the *1*) Poisson, *2*) normalised Poisson and *3*) screened Poisson. Compared to the true distance function, these smooth distance functions achieve varying accuracy properties near and far from the boundary; this is demonstrated in Figure 4.2 for a simple square domain. The varying accuracy properties are essential to determine the required characteristics for robust gait recognition.

4.1.1 Poisson Distance Function

The simplest approach to generate a smooth distance function is based on solving a Dirichlet boundary value problem for the Poisson equation

$$\Delta\varphi = -1 \quad \text{in} \quad \Omega, \qquad \varphi = 0 \quad \text{on} \quad \partial\Omega \tag{4.1}$$

where Ω is a point inside the domain and $\partial\Omega$ represents the boundary which is seen in Figure 4.3. This is a basic mathematical model describing the Brownian motion of particles born at a constant rate inside Ω and dying on $\partial\Omega$. The solution of the Poisson distance function (4.1) is proportional to particle density and is considered a smooth approximation of the true distance function $d(x, \partial\Omega)$.

Compared to the true distance function $d(x, \partial\Omega)$, Figure 4.2 shows the Poisson distance function $\varphi(x)$ achieves a poor approximation near and far from the boundary. The Poisson distance function has been used for action recognition [Gorelick et al. (2007b, 2006)], skeleton extraction [Aubert and Aujol (2012)], turbulence modelling [Tucker (1998)], and geometric de-featuring [Xia et al. (2012)].

4.1.2 Normalised Poisson Distance Function

Normalising the Poisson distance function $\varphi(x)$ can increase the approximation accuracy near $\partial\Omega$. Following Spalding (1994), Tucker (1998) uses

$$\psi(x) = -|\nabla\varphi| + \sqrt{|\nabla\varphi|^2 + 2\varphi} \tag{4.2}$$

which is inspired by the precise distance function reconstruction of 1D cases (4.1) and (4.2). The latter can be re-written

$$\psi(x) = \frac{2\varphi}{\sqrt{|\nabla\varphi|^2 + 2\varphi} + |\nabla\varphi|}$$

and check that

$$\psi = 0 \quad \text{and} \quad \partial\psi/\partial n = 1 \quad \text{on} \quad \partial\Omega \tag{4.3}$$

where n is the inner unit normal to $\partial\Omega$. Therefore the normalised Poisson distance function $\psi(x)$ approximates the true distance function $d(x, \partial\Omega)$ very accurately near $\partial\Omega$. Interestingly, the second normalisation boundary condition in (4.3) has not been noticed before. An alternative normalisation scheme used by geometric modelling [Rvachev (1982),Shapiro (2007)] is

$$\frac{\varphi(x)}{\sqrt{|\nabla\varphi(x)|^2 + \varphi(x)^2}}.$$

To this end, this chapter uses the normalised Poisson distance function $\psi(x)$ based on the combination of (4.1) with (4.2). Compared to the true distance function $d(x, \partial\Omega)$, Figure 4.2 shows the normalised Poisson distance function $\psi(x)$ achieves a poor approximation far from the boundary.

4.1.3 Screened Poisson Distance Function

Finally, the asymptotic relationship between the true distance function and the screened Poisson equation [Varadhan (1967),Theorem 2.3] is exploited. Consider a Dirichlet boundary value problem for a screened Poisson equation in a bounded domain Ω

$$v - t\Delta v = 0 \quad \text{in} \quad \Omega, \qquad v = 1 \quad \text{on} \quad \partial\Omega \tag{4.4}$$

where t is a small, positive parameter. As seen in Varadhan (1967),

$$\lim_{t\to 0} -\sqrt{t}\ln[v(x)] = d(x, \partial\Omega) \tag{4.5}$$

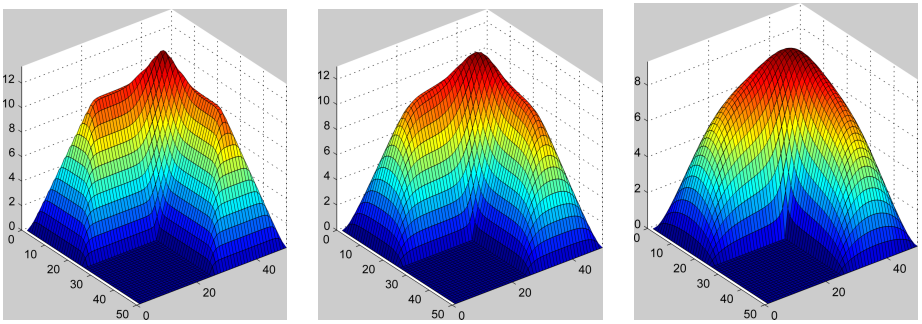
where $d(x, \partial\Omega)$ is the distance from $x \in \Omega$ to $\partial\Omega$, i.e. $d(x, \partial\Omega)$ is approximated by

$$u(x) = -\sqrt{t}\ln v(x) \tag{4.6}$$

which defines a smooth distance function controlled by a smoothing parameter t.

Considering the applications of the screened Poisson distance function, *i*) an inhomogeneous version of (4.4) estimates the distance function from a point set [Gurumoorthy and Rangarajan (2009),Sethi et al. (2012)], *ii*) an anisotropic version of (4.4) traces geodesics on triangulated surfaces [Crane et al. (2013)] and *iii*) (4.6) extracts skeletal structures from greyscale images [Tari et al. (1997)]. Interestingly, the energy in (4.4) is a part of the Ambrosio-Tortorelli elliptic regularisation [Ambrosio and Tortorelli (1990)] of the Mumford-Shah functional [Mumford and Shah (1989)]; further details are found in Shah (1991) and Aubert and Kornprobst (2002) (Section 4.2).

Gurumoorthy and Rangarajan (2009) exploit a variant of the so-called Hopf-Cole



$t = 0.5$ $t = 5$ $t = 50$

Figure 4.4: Screened Poisson distance function (4.6) for an L-shape domain with varying smoothing parameters t. It is clear to see how increasing the smoothing parameter t decreases the approximation accuracy

transformation [Evans (1998)] used in (4.5) where

$$v(x) = \exp\left\{-u(x)/\sqrt{t}\right\} \tag{4.7}$$

is substituted into (4.4) yielding

$$\frac{\partial v}{\partial x_i} = -\frac{v}{\sqrt{t}}\frac{\partial u}{\partial x_i}, \qquad \frac{\partial^2 v}{\partial x_i^2} = \frac{v}{t}\left|\frac{\partial u}{\partial x_i}\right|^2 - \frac{v}{\sqrt{t}}\frac{\partial^2 u}{\partial x_i^2}.$$

Therefore (4.4) is revised

$$0 = v - t\Delta v = v\left[\left(1 - |\nabla u|^2\right) + \sqrt{t}\Delta u\right] \tag{4.8}$$

to provide a regularised eikonal equation for $u(x)$

$$\left(1 - |\nabla u|^2\right) + \sqrt{t}\,\Delta u = 0 \quad \text{in} \quad \Omega, \quad u = 0 \quad \text{on} \quad \partial\Omega. \tag{4.9}$$

To this end, the solution to $u(x)$ (4.9) approximates the true distance function $d(x, \partial\Omega)$ and satisfies the eikonal equation

$$|\nabla d|^2 = 1 \quad \text{in} \quad \Omega, \qquad d = 0 \quad \text{on} \quad \partial\Omega. \tag{4.10}$$

To this end, (4.4) can be computed efficiently with a sparse system of linear equations. Compared to the true distance function $d(x, \partial\Omega)$, Figure 4.2 shows the screened Poisson

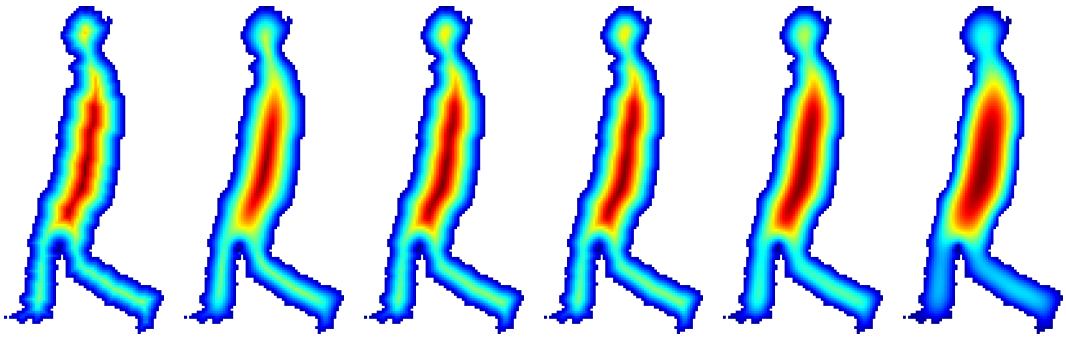
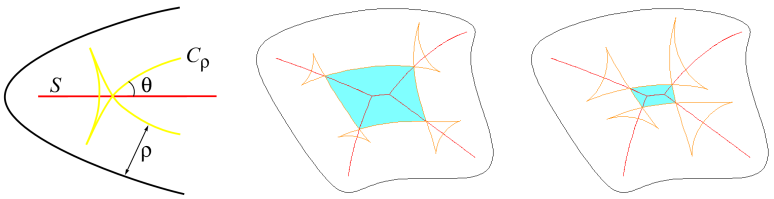


Figure 4.5: Compared to the true distance function, notice how the boundary noise is nicely absorbed by the smooth distance functions

distance function $u(x)$ achieves a poor approximation near $\partial\Omega$; this accuracy property is the opposite of the normalised Poisson distance function $\psi(x)$. The screened Poisson distance function $u(x)$ approximation accuracy is controlled by the smoothing parameter t, and the effects can be seen in Figure 4.4. A small smoothing parameter t yields an accurate approximation of the true distance function $d(x, \partial\Omega)$, while a large t introduces smoothing across the domain thus reducing the approximation accuracy.

Smooth Distance Function Comparison

A visual comparison of the Poisson $\varphi(x)$, normalised Poisson $\psi(x)$ and screened Poisson $u(x)$ distance functions can be seen in Figure 4.5. Compared to the true distance function $d(x, \partial\Omega)$, each smooth distance function yields superior boundary noise absorption capabilities. On top of the varying approximation accuracy, the smooth distance functions vary with respect to smoothing and computational demands. The amount of smoothing incorporated in the smooth distance function is inversely proportional to the approximation accuracy compared to the true distance function. Therefore given a small smoothing parameter t, the screened Poisson distance function achieves the highest approximation accuracy followed by the normalised Poisson and Poisson distance functions. While the computational demands are not a restricting factor in this chapter (real-time gait recognition is not an aim for this thesis), the Poisson distance function achieves the lowest computational complexity followed by the normalised Poisson distance function and screened Poisson distance functions.



(a) (b) (c)

Figure 4.6: (a) illustrates the relationship between skeleton S and offset C_ρ, while (b, c) demonstrate examples of computer generated offsets and the domain skeleton

4.2 Fuzzy Skeletons

By computing the derivative of the smooth distance function, it is possible to extract the fuzzy skeleton. The term fuzzy skeleton is applied due to the thickness compared to the true distance function. Second-order derivatives, such as the Laplacian or curvature, can be used to extract the fuzzy skeleton [Aubert and Aujol (2012),Gorelick et al. (2006),Tari et al. (1997)], however such schemes are more sensitive to noise. Given the noisy nature of gait recognition silhouettes, this chapter uses the squared gradient $|\nabla u|^2$ to extract the fuzzy skeleton.

Consider the observation in Figure 4.6. Assume the boundary $C_0 = \partial\Omega$ of Ω is oriented by its inner normal n, and consider offset curves C_ρ obtained from C_0 by shifting each point of C_0 in the direction of n onto distance ρ. The fuzzy skeleton S of Ω is constructed by the first self-intersections of C_ρ as ρ increases. The self-intersections move along S faster than the offset curves C_ρ move along their normals. If curve C_ρ moves with unit speed, then the self-intersection point moves along S with speed equal to $1/\sin\theta$, where θ is the angle between C_ρ and S. This means the distance function rate of change $d(x, \partial\Omega)$ at that offset self-intersection point $x \in S$ is given by $\sin\theta$. If θ is small at $x \in S$ (and therefore $\sin\theta$ be similarly small) then the orientation normals at the boundary points corresponding to x have near opposite directions and a part of S near x reflects important bilateral symmetry properties of $\partial\Omega$.

The standard 3×3 Sobel kernels are used to compute the gradient of the smooth distance distance function. Taking into account the noisy nature of gait recognition silhouettes, Figure 4.7 demonstrates the fuzzy skeletons extracted by the squared gradient $|\nabla u(x)|^2$ and second-order derivatives. In practice, this means that $|\nabla u(x)|^2$ achieves a fuzzy skeleton which is free from additional spurs.

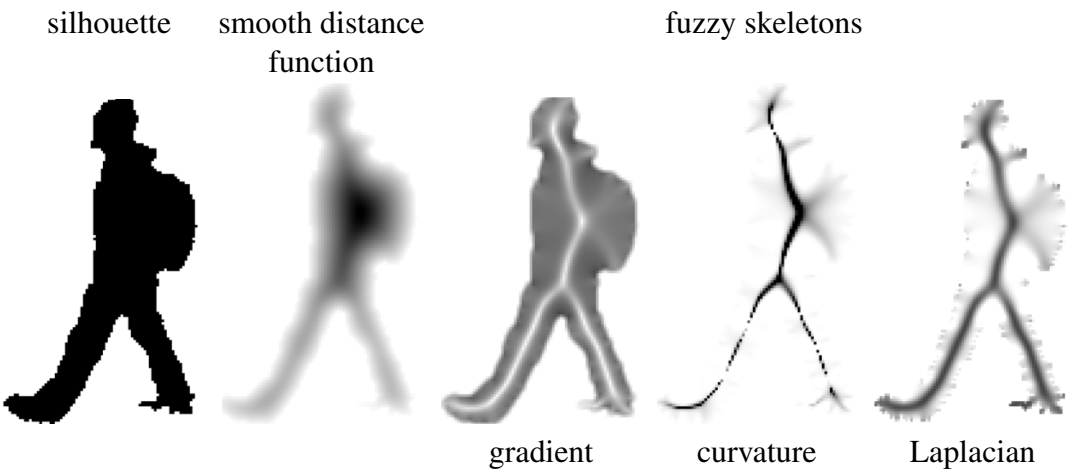


Figure 4.7: Silhouette with corresponding smooth distance function and fuzzy skeletons defined by gradient, curvature, and Laplacian operators; notice that the gradient ensures minimal additional spurs compared to second-order derivatives

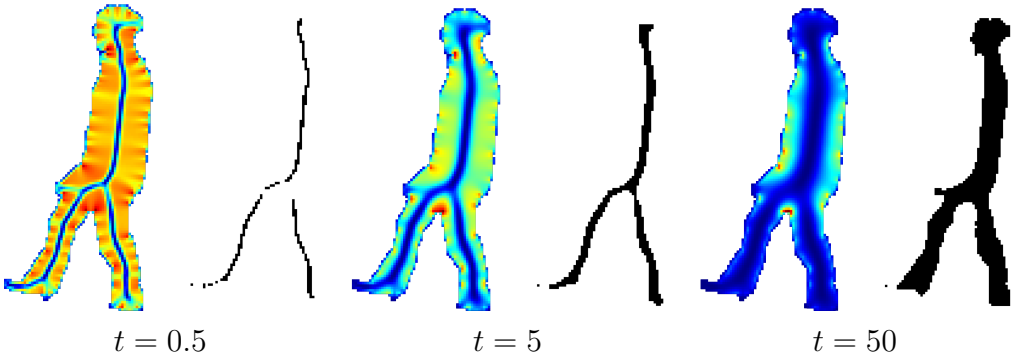
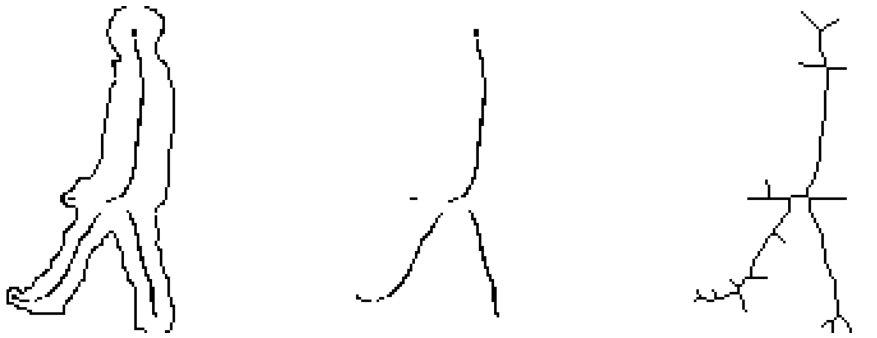


Figure 4.8: Fuzzy skeletons (cooler colours) are extracted when the smooth distance function is convolved with Sobel kernels. For the screened Poisson distance function, the fuzzy skeleton thickness increases with a large smoothing parameter t



unprocessed fuzzy skeleton fuzzy skeleton true skeleton

Figure 4.9: Low gradient magnitude values correspond to the unprocessed fuzzy skeleton and silhouette boundary (left). The fuzzy skeleton (middle) is be extracted by removing a small number of boundary layers. Compared to the true skeleton extracted via the medial axis transform (right), the fuzzy skeleton is considerably smoother and contains significantly fewer spurs

Given the screened Poisson distance function $u(x)$ is dependent on smoothing parameter t, Figure 4.8 demonstrates the fuzzy skeletons computed from the squared gradient $|\nabla u(x)|^2$. While a small smoothing parameter t yields an accurate approximation of the true distance function $d(x, \partial\Omega)$, in practice this yields an unstable, discontinuous fuzzy skeleton; this might be rectified through Canny's hysteresis thresholding procedure [Canny (1986)]. However a larger smoothing parameter t yields a lower accuracy approximation and a thicker fuzzy skeleton which begins to resemble the original silhouette.

Computing the magnitude of the squared gradient $|\nabla u(x)|^2$ exposes the fuzzy skeleton seen in Figure 4.9. However in practice this also extracts the silhouette boundary. To resolve this issue, a small number of boundary layers are removed to yield the fuzzy skeleton. Compared to the true skeleton, it is clear to see that the fuzzy skeleton is free from additional spurs introduced by boundary noise.

4.3 Skeleton Variance Image

The Skeleton Variance Image (SVIM) is the novel combination of a fuzzy skeleton representation and single compact 2D gait representation. Similar to the Gait Energy Image, the SVIM requires space-normalisation and time-normalisation which are standard single compact 2D representation procedures.

Space-normalisation requires size-normalisation and horizontal alignment. Given the process of extracting fuzzy skeletons relies on variables which can alter the reference

Figure 4.10: SVIM for TUM GAID covariate factors normal (N), bag (B), shoes (S), time and normal (TN), time and bag (TB) and time and shoes (TS)

point for horizontal alignment, it is essential that space-normalisation is performed on the silhouette sequence. Finally, time-normalisation condenses the fuzzy skeleton sequence into a single compact 2D gait representation by computing the pixel-wise variance (5.1) of the fuzzy skeleton sequence by

$$SVIM(x,y) = \frac{\left(S_m - \left(\frac{1}{N}\sum_{m=1}^{N} S_m(x,y)\right)\right)^2}{N-1} \tag{4.11}$$

where N is the number of fuzzy skeletons in the sequence, m is the fuzzy skeleton number, x and y are the 2D spatial image coordinates and S is a fuzzy skeleton. In practice, (4.11) means the SVIM expresses motion features which are consistent over time and achieve reduced sensitivity to covariate factors.

4.4 Experimental Procedure

4.4.1 SVIM Construction

The SVIM requires space-normalisation prior to (4.11). This is achieved by horizontally aligning the centroid of the top 10% of the silhouette (head) as a reference point. This process is similar to Hofmann et al. (2011) and the same as Section 3.3.2.1.

4.4.2 Dataset

The SVIM is validated in the CASIA B and TUM GAID datasets which are explained in Chapter 2.4.

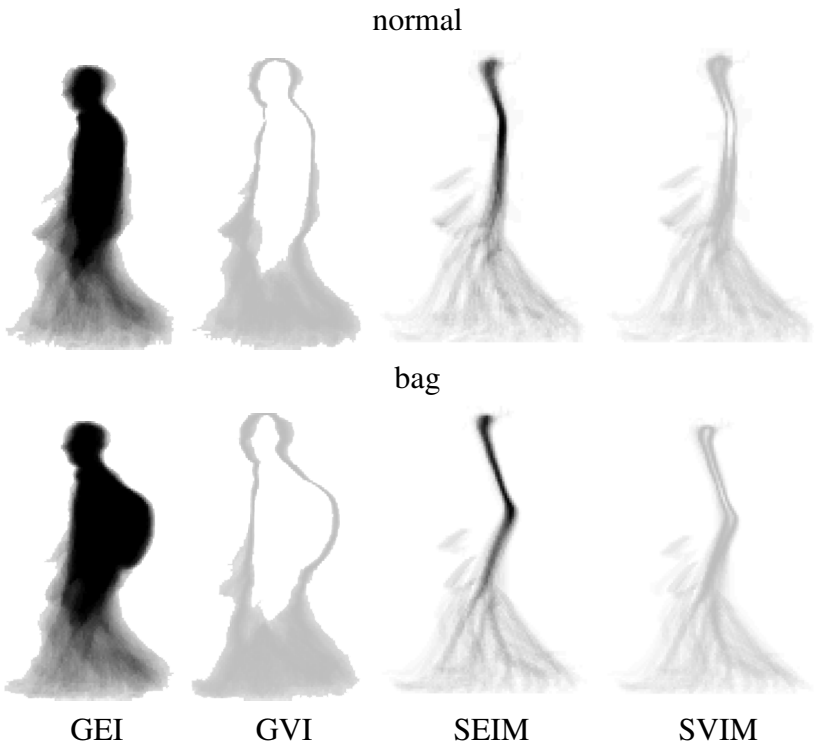


Figure 4.11: Gait Energy Image (GEI), Gait Variance Image (GVI), Skeleton Energy Image (SEIM) and Skeleton Variance Image (SVIM). Notice how the carried bag covariate factor and its motion is encoded

4.4.3 Baseline and Comparable Representations

The following baseline and comparable gait representations are seen in Figure 4.11. The Gait Energy Image (GEI) [Han and Bhanu (2006)], seen in Figure 4.11, is the performance baseline for the SVIM. The GEI is computed using

$$GEI(x,y) = \frac{1}{N}\sum_{m=1}^{N} B_m(x,y) \tag{4.12}$$

where N is the number of silhouettes in the sequence, m is the silhouette number, x and y are the 2D spatial image coordinates and B is a silhouette.

To further examine the performance of motion features and fuzzy skeletons, two further novel representations are defined. The Gait Variance Image (GVI), seen in Figure 4.11, is analogous to the SVIM where the silhouette sequence replaces the fuzzy skeleton sequence

$$GVI(x,y) = \frac{\left(B_m - \left(\frac{1}{N}\sum_{m=1}^{N} B(x,y)\right)\right)^2}{N-1} \tag{4.13}$$

where N is the number of silhouettes in the sequence, m is the silhouette number, x and y are the 2D spatial image coordinates and B is a silhouette. Finally the Skeleton Energy Image (SEIM), seen in Figure 4.11, is analogous to the GEI where the fuzzy skeleton sequence replaces the silhouette sequence

$$SEIM(x,y) = \frac{1}{N}\sum_{m=1}^{N} S_m(x,y) \tag{4.14}$$

where N is the number of fuzzy skeletons in the sequence, m is the fuzzy skeleton number, x and y are the 2D spatial image coordinates and S is a skeleton.

These gait representations allow a performance comparison of *a*) appearance and motion features (GEI and SEIM) and motion features (GVI and SVIM) and *b*) silhouette representations (GEI and GVI) and skeleton representations (SEIM and SVIM).

4.4.4 Smooth Distance Function

The Poisson, normalised Poisson and screened Poisson distance functions are evaluated to determine their contribution to SVIM performance, For the screened Poisson distance function, a broad range of smoothing parameters $\{t = 0.1, 0.5, 5, 10...$ in steps of 10 to $90\}$ are analysed given the impact on approximation accuracy (Figure 4.4) and skeleton thickness (Figure 4.8). To recap, a small smoothing parameter t yields an accurate approximation of the true distance function and a thin (and often unstable) fuzzy skeleton. A large smoothing parameter t yields a poorer approximation of the true distance function and a thick fuzzy skeleton. Overall, each smooth distance function achieves varying accuracy properties near and far from the silhouette boundary; this is essential to determine the requirements for robust gait recognition.

4.4.5 Dimensionality Reduction and Classification

The following dimensionality reduction and classification procedures are standard [Han and Bhanu (2006)] for single compact 2D gait representations. In particular, these procedures are effective for the small number of training sequences in gait recognition datasets. The GEI, GVI, SEIM and SVIM (standard dataset image sizes - CASIA B: 240×240,

CASIA B dataset (124 test persons)

Features	Gait Representation (%)	nm	bg	cl	average
Appearance and motion	GEI (baseline)	**100.0**	53.2	22.2	58.5
	SEIM (*Poisson*)	99.2	83.1	60.9	81.0
	SEIM (*normalised Poisson*)	98.4	84.7	61.3	81.5
	SEIM (*screened Poisson*)	98.0	**93.1**	69.8	87.0
Motion	GVI	97.2	77.8	50.4	75.1
	SVIM (*Poisson*)	98.4	92.3	67.3	86.0
	SVIM (*normalised Poisson*)	98.8	80.2	65.7	81.6
	SVIM (*screened Poisson*)	98.4	92.7	**71.8**	**87.6**

TUM GAID dataset (155 test persons)

Features	Gait Representation (%)	N	B	S	TN	TB	TS	average
Appearance and motion	GEI (baseline)	**99.7**	19.0	**96.5**	34.4	0.0	43.8	67.4
	SEIM (*Poisson*)	97.4	8.1	89.7	40.6	3.1	28.1	61.2
	SEIM (*normalised Poisson*)	99.0	18.4	96.1	15.6	3.1	28.1	66.0
	SEIM (*screened Poisson*)	98.7	18.4	96.1	31.3	0.0	31.3	66.4
Motion	GVI	99.0	47.7	94.5	62.5	15.6	**62.5**	77.3
	SVIM (*Poisson*)	97.4	53.6	88.1	**65.6**	21.9	53.1	76.6
	SVIM (*normalised Poisson*)	98.4	54.2	92.9	50.0	28.1	37.5	77.8
	SVIM (*screened Poisson*)	98.4	**64.2**	91.6	**65.6**	**31.3**	50.0	**81.4**

Table 4.1: CASIA B and TUM GAID dataset results for representations: GEI, SEIM, GVI and SVIM, and sequences: normal (nm, N), carrying a bag (bg, B), clothing (cl), shoes (S), time and normal (TN), time and bag (TB), time and shoes (TS); distance functions are based on the Poisson, normalised Poisson and screened Poisson (for the smoothing parameter t achieving the highest average performance across covariate factors)

TUM GAID: 128×178) are reshaped into a 1D feature vector (CASIA B: $57600D$, TUM GAID: $22784D$) to describe gait. To reduce the high dimensionality, Principle Component Analysis and Linear Discriminant Analysis (code adapted from [van der Maaten (2007)]) are combined to satisfy the best data representation with respect to covariance and class separability respectively (CASIA B: $123D$, TUM GAID: $154D$ account for $\approx$ 97% variance). Nearest Neighbour classification is performed using the Euclidean and Cosine distance which are standards defined by the CASIA B and TUM GAID datasets respectively. The SVIM results are compared to the ground truth in a confusion matrix. The average of the diagonal in the confusion matrix yields the correct classification.

4.5 Results and Discussion

The results for each representation (SVIM, SEIM, GVI, GEI) and smooth distance function (Poisson, normalised Poisson, screened Poisson) are presented in Table 4.1. The screened Poisson smoothing parameter t is analysed in Figure 4.12 and Figure 4.13 for the CASIA B and TUM GAID datasets respectively. Given these results, the performance is discussed based on: *a*) covariate factor performance trends, *b*) appearance and motion features versus motion features, *c*) silhouette representations versus fuzzy skeleton representations, *d*) smooth distance functions, *e*) smoothing parameter t and *f*) general recommendations.

4.5.1 Covariate Factor Performance Trends

The following observations concerning Table 4.1 occur regardless of representation (GEI, GVI, SEIM, SVIM) and dataset (CASIA B and TUM GAID).

The covariate factor free sequences (CASIA B: nm, TUM GAID: N, S) achieve a high performance due to their visual similarities with training sequences; this demonstrates the proof of concept. The shoe sequences involve shoe covers which incur negligible appearance and motion alterations from training sequences and thus achieve a high performance; note that the covariate factor is aimed towards acoustic gait recognition (identifying a person based on the sounds made while walking) given the considerable acoustic impact of the shoe covers rustling. However shoe types such as heels or flip flops [Bouchrika and Nixon (2008)] can significantly affect the natural appearance and motion of gait. Compared to Chapter 3 (HOG describing the GEI), it is interesting to note that the SVIM achieves superior shoe sequence performance; this is attributed to the discriminative features extracted from the GEI, GVI, SEIM and SVIM representations.

The covariate factor sequence (CASIA B: bg, cl, TUM GAID: B, TN, TB, TS) performance is generally poorer given the unique ways in which covariate factors affect the natural appearance and motion of gait; this subsequently reduces the visual similarities to training sequences. Bag sequences occur in both datasets where performance differences can be partly explained by the varying bag type; the CASIA B dataset uses varying bag types (rucksacks, handbags etc. carried in varying locations) while the TUM GAID

dataset uses a consistent rucksack. Clothing sequences in the CASIA B dataset are challenging given the varying jacket shapes and lengths. The complex coupled time-based covariate factors in the TUM GAID dataset incur significant performance drops, in some cases the performance can be half that achieved with single covariate factors; see Matovski et al. (2012) for further information regarding time as a covariate factor. This occurs as coupled covariate factors can significantly affect the natural appearance and motion of gait compared to single covariate factors.

4.5.2 Appearance and Motion Features versus Motion Features

Appearance and motion features are extracted with (4.12) and (4.14) which are effective for noise mitigation. Motion features are extracted with (4.11) and (4.13) which risk amplifying noise. Given the higher quality silhouettes in the TUM GAID dataset, Table 4.1 shows that extracting motion features can achieve approximately double the performance achieved when extracting appearance and motion features. However the performance difference is negligible in the CASIA B dataset due to the poorer quality silhouettes. Using appearance features can be unreliable given their sensitivity to covariate factors. Conversely, motion features are effective given their consistency over time and reduced sensitivity to covariate factors.

4.5.3 Silhouette versus Fuzzy Skeleton Representations

Silhouettes represent covariate factors as a mass of pixels, while covariate factors manifest themselves as a bend in the fuzzy skeleton. Due to the poorer quality silhouettes in the CASIA B dataset, there is a considerable performance difference between fuzzy skeleton representations and silhouette representations. However the higher quality silhouettes in the TUM GAID dataset show the performance difference is negligible. Depending on the smooth distance function, Table 4.1 shows that fuzzy skeleton representations outperform silhouette representations during covariate factor sequences. This occurs as fuzzy skeletons address an existing limitation of gait recognition i.e. emphasizing gait motion whilst suppressing covariate factor motion.

4.5.4 Smooth Distance Function

Table 4.1 demonstrates that the screened Poisson distance function is superior regardless of dataset and representation (GEI, GVI, SEIM, SVIM); the normalised Poisson and Poisson distance functions follow thereafter. The screened Poisson distance function is beneficial for covariate factor sequences due to the tunable smoothing parameter t. Overall, Table 4.1 shows that covariate factor free sequences are relatively insensitive to distance function approximation accuracy. Conversely, covariate factor sequences are sensitive to distance function approximation accuracy.

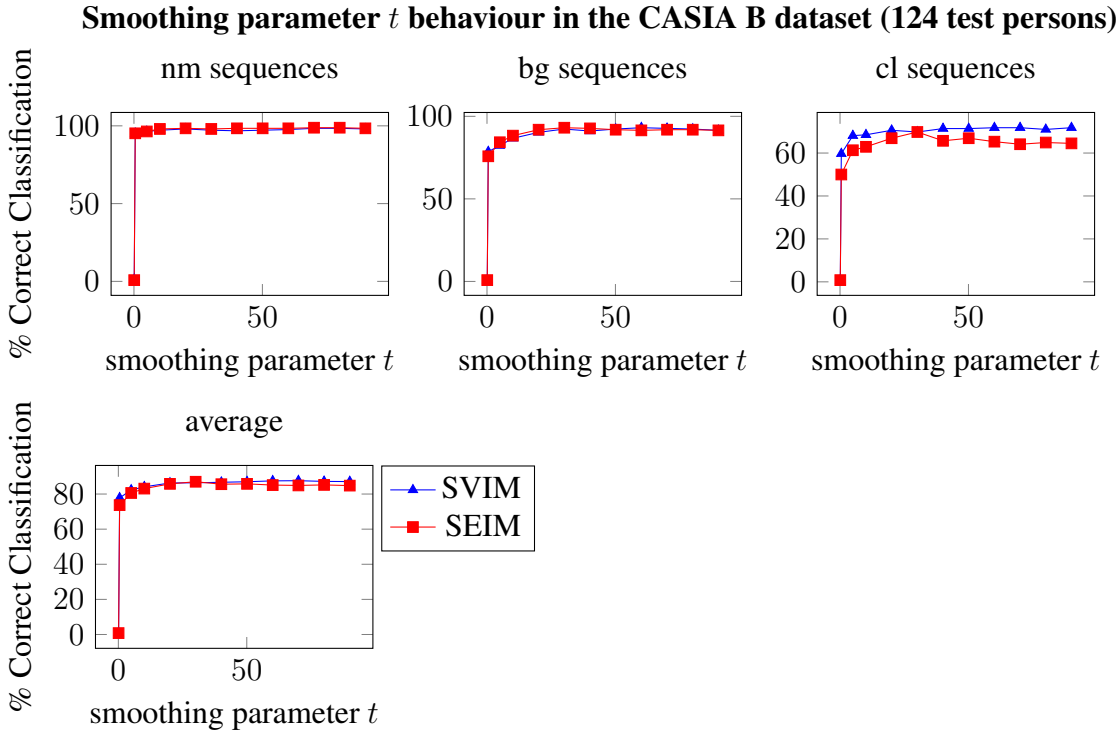


Figure 4.12: SVIM and SEIM CASIA B dataset performance when using the screened Poisson distance function with respect to smoothing parameter t: normal (nm), carrying a bag (bg), clothing (cl), average. A small t yields a poor performance due to the unstable skeleton

4.5.5 Smoothing Parameter t Behaviour

The smoothing parameter t is responsible for the approximation accuracy (Figure 4.4) and fuzzy skeleton thickness (Figure 4.8). The performance of the SEIM and SVIM with smoothing parameter t is seen in Figure 4.12 and Figure 4.13 for the CASIA B and TUM GAID datasets respectively.

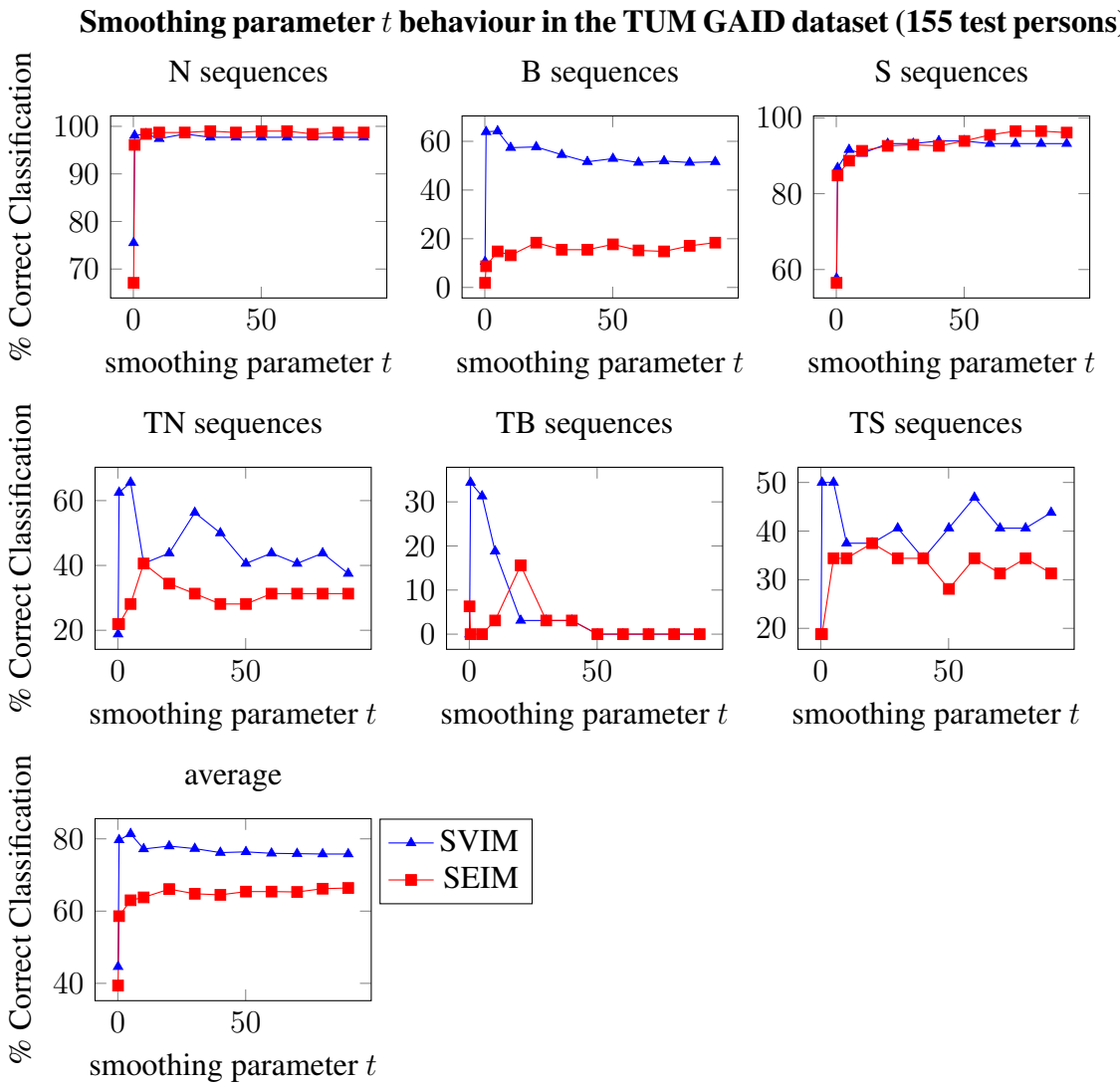


Figure 4.13: SVIM and SEIM TUM GAID dataset performance when using the screened Poisson distance function with respect to smoothing parameter t: normal (N), carrying a bag (B), shoes (S), time and normal (TN), time and carrying a bag (TB), time and shoes (TS), average. A small t yields a poor performance due to the unstable skeleton

Consider the CASIA B dataset results in Figure 4.12 where the performance patterns are repetitive across all sequences. While a small smoothing parameter ($t = 0.5$) achieves a good approximation of the true distance function, in practice this yields an unstable fuzzy skeleton and thus poor performance. The performance begins to plateau with larger smoothing parameters ($t \geq 5$). Overall, the average performance across all sequences show the SVIM marginally outranks the SEIM due to the poor silhouette quality.

Now consider the TUM GAID dataset results in Figure 4.13. Single covariate factor sequences (N, B, S) mimic similar performance patterns to the CASIA B dataset (i.e. a small smoothing parameter t yields unstable fuzzy skeletons and poor performance). The time-based covariate factor sequences (TN, TB, TS) following similar performance

patterns with greater fluctuations due to the coupled covariate factors. Overall, the average performance across the sequences show the SVIM outranks the SEIM due to the higher silhouette quality.

To achieve the highest average performance across covariate factors, the CASIA B and TUM GAID datasets prefer $t = 70$ and $t = 5$ for the SVIM respectively, and $t = 30$ and $t = 90$ for the SEIM respectively. This indicates there is no single optimised smoothing parameter t given the sensitivity to *i*) covariate factors due to the unique way in which the natural appearance and motion of gait are affected, and *ii*) the silhouette quality where boundary noise and missing heads or limbs cause visual dissimilarities between training and test skeletons. Across the datasets, a larger smoothing parameter t is required for robust gait recognition. This indicates that a high accuracy smooth distance function is not necessary due to the quantity of smoothing required to absorb the boundary noise.

The difference in smoothing parameter may be explained by the difference in image size (CASIA B: 240×240, TUM GAID: 128×178). If the image Ω is scaled by a factor s whilst keeping a fixed resolution and assuming the solution $v(x)$ to (4.4) remains invariant, this leads to scaling the smoothing parameter t by s^2. It should also be noted that the smoothing parameter t may be sensitive to application.

4.5.6 General Recommendations

The results in this chapter indicate that the SVIM is superior for gait recognition, i.e. motion features extracted from a fuzzy skeleton derived from the screened Poisson distance function. This is due to three primary factors *a*) the SVIM extracts discriminative motion features which are more consistent over time and less sensitive to covariate factors, *b*) covariate factors manifest themselves as a bend in the fuzzy skeleton which therefore mitigates covariate factor motion and addresses a limitation of gait recognition, and *c*) the tunable smoothing parameter t which is effective for covariate factor generalisation.

CASIA B dataset dataset (124 test persons)

Gait Representation (%)	nm	bg	cl	average
CGI Wang et al. (2012)	88.1	43.7	43.0	58.2
GEI Han and Bhanu (2006)	100.0	53.2	22.2	58.5
GEnI Bashir et al. (2010)	**100.0**	78.3	44.0	74.1
MII + MDIs Bashir et al. (2009b)	97.5	83.6	48.8	76.6
SEI + GSP Huang and Boulgouris (2012)	99.0	64.0	72.0	78.3
P_{RW}GEI Yogarajah et al. (2011)	98.4	**93.1**	44.4	78.6
Body segmentation Li et al. (2010)	99.2	80.6	75.8	85.2
AEI Zhang et al. (2010)	98.4	91.9	72.2	87.5
SGEI + GEI Li and Chen (2013)	98.2	80.7	**83.9**	87.6
M_G Bashir et al. (2008a)	**100.0**	91.0	80.6	**90.5**
SVIM (proposed)	98.4	92.7	71.8	87.6

TUM GAID dataset (155 test persons)

Gait Representation (%)	N	B	S	TN	TB	TS	average
depth GEI Han and Bhanu (2006)	**99.7**	17.4	**96.5**	37.5	0.0	43.8	67.1
GEV Hofmann et al. (2013)	94.2	13.9	87.7	41.0	0.0	31.0	61.4
DGHEI Hofmann et al. (2013)	99.0	40.3	96.1	50.0	0.0	44.0	74.1
SVIM (proposed)	98.4	**64.2**	91.6	**65.6**	**31.3**	**50.0**	**81.4**

Table 4.2: Comparison of the SVIM results to current state-of-the-art results in the CASIA B and TUM GAID datasets for normal (nm, N), carrying a bag (bg, B), clothing (cl), shoes (S), time and normal (TN), time and bag (TB), time and shoes (TS). The SVIM significantly increases TUM GAID dataset state-of-the-art performance by 9.9%, whilst ranking second in the CASIA B dataset

4.6 Comparison to State-of-the-Art

The SVIM is compared to current state-of-the-art results for the CASIA B and TUM GAID datasets in Table 4.2. SVIM results are based on the screened Poisson distance function and the highest performing smoothing parameter t in each dataset.

While the SVIM ranks second in the CASIA B dataset, the SVIM achieves a 9.9% increase over current state-of-the-art results in the TUM GAID dataset. State-of-the-art results may not be achieved using the CASIA B dataset due to the poorer silhouette quality (compared to the TUM GAID dataset). This reduces the visual similarities between training and test SVIMs and demonstrates the sensitivity to silhouette quality. It is interesting to note that the SVIM only benefits the covariate factor sequences (CASIA B: bg, cl, TUM GAID: B, TN, TB, TS) while incurring a minor performance drop during covariate factor free sequences (CASIA B: nm, TUM GAID: N, S). This occurs as covariate factor generalisation is paramount meaning a non-optimal smoothing parameter t is used for covariate factor free sequences.

The SVIM is successful due to *i*) the extraction of discriminative motion features and *ii*) the fuzzy skeleton emphasizing gait motion and mitigating covariate factor motion. Overall, the SVIM achieves an important goal of promoting robust and tunable skeleton gait representations which alleviate the common limitation of boundary noise sensitivity.

4.7 Conclusion

The success of the SVIM is due to three primary reasons.

1. Skeletons are infrequently used in gait recognition due to their sensitivity to boundary noise. Therefore a smooth distance function derived from the screened Poisson equation is used to absorb the boundary noise. The screened Poisson distance function is beneficial due to the tunable smoothing parameter t which is capable of achieving superior covariate factor generalisation.

2. While covariate factors are static with respect to the human body, covariate factors naturally undergo motion due to the nature of human gait. Given covariate factors

manifest themselves as bends in the fuzzy skeleton, the SVIM can emphasize gait motion whilst suppressing covariate factor motion; this addresses a current limitation in gait recognition research.

3. The SVIM is based on motion features which are more consistent over time compared to appearance features. This means that the SVIM is effective during the complex coupled time-based sequences in the TUM GAID dataset.

Hypothesis Revised

At the beginning of this chapter, the following hypothesis was made:

> "This chapter argues that by exploiting the Poisson equation to construct a smooth distance function, fuzzy skeletons can be extracted and formed into a single compact 2D gait representation to yield a discriminative gait descriptor."

Therefore during the course of this chapter, the hypothesis has been verified.

Future Directions

Further developments could be made to increase the robustness of the SVIM and also widen its application.

- For the SVIM to be used in another dataset or application, it would be beneficial to promote an optimal smoothing parameter t which is invariant to image size.
- There is no smoothing parameter t capable of generalising over all covariate factors. This is natural given the unique way in which covariate factors affect the natural appearance and motion of gait. However, if covariate factors could be detected and recognised, the optimal smoothing parameter t could be applied. Given silhouettes reject all colour and texture cues, covariate factor detection and recognition would be best achieved using RGB images.

- Suppressing the influence of covariate factors is paramount for gait recognition. Achieving this at the silhouette extraction stage would be beneficial, e.g. distance-based shape priors for image segmentation [Cremers (2013)].

While the SVIM achieves significant state-of-the-art advances, a small degree of covariate factor artefacts remain. This is natural given silhouettes reject all colour and texture cues meaning it is impossible to remove every covariate factor related pixel. Therefore Chapter 5 is devoted to detecting and removing covariate factors in single compact 2D gait representations.

Chapter 5

Covariate Factor Detection & Removal

This chapter is devoted to detecting and removing covariate factors in single compact 2D gait representations. Existing covariate factor detection techniques cannot effectively differentiate between natural gait motion and covariate factor motion; this means removal techniques retain covariate factor artefacts. Therefore this limitation is addressed by establishing the pixel-wise composition of covariate factors. The novel covariate factor detection and removal (CFDR) module trials a) four detection techniques for their ability to differentiate between gait motion and covariate factor motion and b) three removal techniques for their ability to effectively remove covariate factors. Across validation datasets, single compact 2D gait representations with the CFDR module applied achieve a 15.3% performance increase.

Hypothesis

This chapter argues that single compact 2D gait representations can achieve superior robustness when performing dedicated covariate factor detection and removal.

Publications

The results of this chapter have been presented at the International Symposium on Visual Computing [Whytock et al. (2013c)] and the International Conference on Imaging for Crime Detection and Prevention [Whytock et al. (2013b)], and published in the Journal of Machine Vision and Applications [Whytock et al. (2015)].

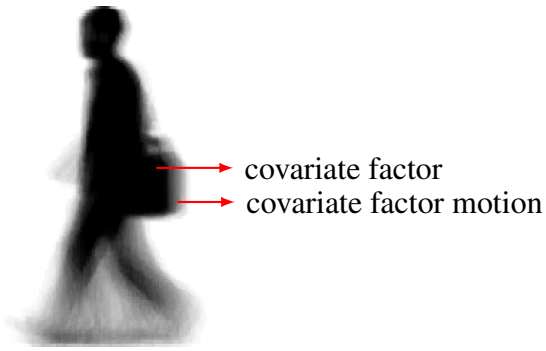


Figure 5.1: Existing covariate factor detection techniques rely on a simplifying assumption which states covariate factors are static with respect to the body. However the nature of human gait causes covariate factors to subsequently undergo motion. This observation is demonstrated with the GEI but occurs regardless of representation

Robust gait recognition is achieved by mitigating the effects of covariate factors. For single compact 2D gait representations, existing covariate factor detection and removal techniques are designed for a given representation and rely on a simplifying assumption: covariate factors are static with respect to the body. While this is somewhat true, covariate factors subsequently undergo motion due to the natural motion of gait; this is demonstrated by the Gait Energy Image (GEI) in Figure 5.1, however the observation is true regardless of single compact 2D gait representation. This causes pixel-wise similarities between natural gait motion and covariate factor motion meaning removal techniques can only partially mitigate the effects of covariate factors.

There are three notable covariate factor detection and removal techniques which are based on the Gait Energy Image (GEI). Bashir et al. (2008a) use a pixel intensity threshold which defines covariate factors as static features. The corresponding pixel locations are used to form a removal mask. A secondary removal mask is based on all pixel values from the upper two thirds of the GEI. The covariate factors are removed by applying the M_G mask which is formed by the binary 'AND' operation of both removal masks. This is an effective combination of covariate factor detection and removal, however the M_G mask is limited by the assumption about the pixel-wise composition of covariate factors. Li et al. (2010) use a threshold to extract the static features from the GEI which are segmented into six sections using anthropometrics. A threshold for pixel distribution is determined for each section based on the the normal (covariate factor free) GEI. Any section whose pixel distribution exceeds the threshold is removed from the GEI. This covariate factor

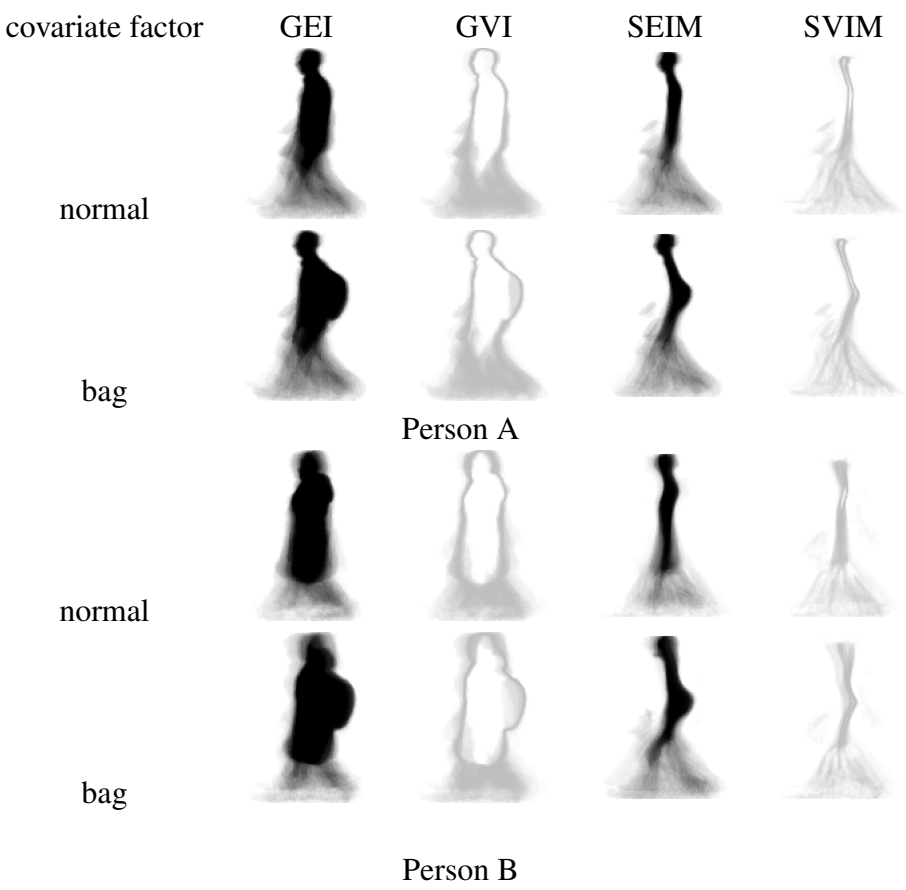


Figure 5.2: GEI, GVI, SEIM and SVIMs from two persons in the TUM GAID dataset for normal (covariate factor free) and bag sequences; notice how each representation maintains the unique nature of gait, and how covariate factors are encoded by silhouettes and skeletons

detection and removal approach is limited by *a*) the assumption about covariate factor pixel composition and *b*) that whole sections can be removed even if a covariate factor partially occupies the section. Huang and Boulgouris (2012) use anthropometrics to segment the body into sections containing the head, torso and legs. The horizontal centres of each section are aligned in order to mitigate the effects of body rotations caused by covariate factors. This approach is limited by the suppression of natural body rotations.

To this end, this chapter is devoted to *a*) understanding the pixel-wise composition of covariate factors, *b*) retaining the natural inter-class and intra-class variance of human gait and *c*) forming a generalised covariate factor detection and removal (CFDR) module for single compact 2D gait representations.

CASIA B dataset (124 test persons)

Gait Representation (%)	nm	bg	cl	average
GEI	**100.0**	53.2	22.2	58.5
GVI	97.2	77.8	50.4	75.1
SEIM	98.0	**93.1**	69.8	87.0
SVIM	98.4	92.7	**71.8**	**87.6**

TUM GAID dataset (155 test persons)

Gait Representation (%)	N	B	S	TN	TB	TS	average
depth GEI	**99.7**	19.0	**96.5**	34.4	0.0	43.8	67.4
GVI	99.0	47.7	94.5	62.5	15.6	**62.5**	77.3
SEIM	98.7	18.4	96.1	31.3	0.0	31.3	66.4
SVIM	98.4	**64.2**	91.6	**65.6**	**31.3**	50.0	**81.4**

Table 5.1: GEI, GVI, SEIM and SVIM results for the CASIA B and TUM GAID datasets; these are baseline values before the CFDR module is applied. CASIA B dataset sequences: normal (nm), carrying a bag (bg), clothing (cl); TUM GAID dataset sequences: normal (N), carrying a bag (B), shoes (S), time and normal (TN), time and carrying a bag (TB), time and shoes (TS)

5.1 Validation Gait Representations

Single compact 2D gait representations are highly cited and frequently employed in gait recognition due to their *i*) robustness to noise and short-term occlusions, *ii*) reduced computational demands and *iii*) discriminative features relating to the appearance (pose) and motion of gait. The CFDR module is applied to the Gait Energy Image (GEI), Gait Variance Image (GVI), Skeleton Energy Image (SEIM) and Skeleton Variance Image (SVIM). These representations are seen in Figure 5.2 and Table 5.1 shows the baseline (without the CFDR module applied) results for the CASIA B and TUM GAID dataset. The CFDR module has the potential to enhance the robustness of the GEI, GVI, SEIM and SVIM by *a*) maximising covariate factor detection, *b*) minimising the pixel-wise confusion between natural gait motion and covariate factor motion and *c*) retaining the natural inter-class and intra-class variance in human gait.

Gait Energy Image The GEI [Han and Bhanu (2006)], seen in Figure 5.2, uses

$$GEI(x,y) = \frac{1}{N}\sum_{m=1}^{N} B_m(x,y) \tag{5.1}$$

where N is the number of silhouettes in the sequence, m is the silhouette number, x and y are the 2D spatial image coordinates and B is a silhouette. The GEI shows high and low pixel intensity values corresponding to static (pose) features and dynamic (motion) features respectively. Due to covariate factor motion, covariate factors are seen as a core of static features surrounded by dynamic features. The GEI is limited by the *a*) pixel-wise confusion between natural gait motion and covariate factor motion and *b*) the sensitivity of static features to covariate factors.

Gait Variance Image The GVI [Whytock et al. (2015)], seen in Figure 5.2, uses

$$GVI(x,y) = \frac{\left(B_m - \left(\frac{1}{N}\sum_{m=1}^{N} B_m(x,y)\right)\right)^2}{N-1} \tag{5.2}$$

where N is the number of silhouettes in the sequence, m is the silhouette number, x and y are the 2D spatial image coordinates and B is a silhouette. The GVI shows low pixel intensity values corresponding to dynamic features. For covariate factor sequences, Table 5.1 shows that the GVI is beneficial given the extraction of dynamic features which are less sensitive to covariate factors. However, the GVI is limited by the pixel-wise confusion between natural gait motion and covariate factor motion.

Skeleton Energy Image and Skeleton Variance Image The SEIM [Whytock et al. (2015)] uses

$$SEIM(x,y) = \frac{1}{N}\sum_{m=1}^{N} S_m(x,y) \tag{5.3}$$

while the SVIM [Whytock et al. (2015)] uses

$$SVIM(x,y) = \frac{\left(S_m - \left(\frac{1}{N}\sum_{m=1}^{N} S_m(x,y)\right)\right)^2}{N-1} \tag{5.4}$$

Figure 5.3: Detecting the extent of the carried bag covariate factor is considerably simpler in an RGB image compared to a silhouette figure

where N is the number of fuzzy skeletons in the sequence, m is the fuzzy skeleton number, x and y are the 2D spatial image coordinates and S is the fuzzy skeleton. Figure 5.2 shows covariate factors appear as a bend in the fuzzy skeleton. This helps the SEIM and SVIM mitigate covariate factors where the SVIM achieves a higher performance in Table 5.1. Due to the extraction of static features and dynamic features, the limitations of the GEI apply to the SEIM; similarly, the limitations of the GVI apply to the SVIM given the extraction of dynamic features.

5.2 Covariate Factor Detection

The GEI, GVI, SEIM and SVIM are founded on silhouettes which reject colour and texture cues to avoid bias to gait appearance given the sensitivity to time. Consider Figure 5.3 which shows a person carrying a bag in an RGB image and the equivalent silhouette figure. By human eye, detecting the carried bag in the RGB image is straightforward. However it is impossible to detect every covariate factor related pixel value in the silhouette figure as the extent to which the bag encroaches the silhouette is unknown.

Detecting covariate factors in a gait representation (GR, i.e. the GEI, GVI, SEIM, SVIM) requires a three stage process: *1*) construct the "typical" GR to establish the pixel intensity distribution of covariate factor free (normal) sequences, *2*) apply a tolerance to the "typical" GR to retain the natural inter-class and intra-class variance in human gait and finally *3*) compare the "typical" GR to a test GR to detect covariate factors.

Figure 5.4: "Typical" representations tGR (top) versus test representations (bottom); notice the smoothing around the lower limb area

Stage 1: "Typical" GR

The "typical" GR (tGR) is used to determine how the body is posed and distributed (with respect to pixel intensity values) in each GR. Given gait recognition datasets provide standardised training sequences, this can be achieved by space-normalising and time-normalising the corresponding GRs. Therefore tGR is defined as

$$tGR(x, y) = \frac{1}{N} \sum_{m=1}^{N} GR_m(x, y) \tag{5.5}$$

where N is the number of training GRs, m is the training GR number, x and y are the 2D spatial image coordinates and GR is a training GR. The averaging in (5.5) causes dynamic feature smoothing which is seen in Figure 5.4.

Stage 2: "Typical" GR Tolerance

The tGR requires a degree of tolerance to incorporate the natural intra-class and intra-class variance in human gait which may have been mitigated by Stage 1. The tolerance is given by

$$\sigma(x, y) = \sqrt{\frac{1}{N} \sum_{m=1}^{N} (GR_m(x, y) - tGR(x, y))^2} \tag{5.6}$$

where N is the number of training GRs, m is the training GR number, x and y are the 2D spatial image coordinates, GR is a training GR and tGR is the "typical" GR. Four levels of tolerance are applied to tGR

$$\begin{aligned} tGR_0(x,y) = tGR(x,y) \quad tGR_1(x,y) = tGR(x,y) \pm \sigma(x,y) \\ tGR_2(x,y) = tGR(x,y) \pm 2\sigma(x,y) \quad tGR_3(x,y) = tGR(x,y) \pm 3\sigma(x,y) \end{aligned} \tag{5.7}$$

where $tGR_{0,1,2,3}$ is the "typical" GR with the relevant tolerance $\{_{0,1,2,3}\}$ applied, x and y are the 2D spatial image coordinates, tGR is the "typical" GR and σ is the standard deviation of all training GRs. Tolerance is required either side of tGR due to the uniqueness of gait cause by the natural inter-class and intra-class variation of human gait. Note that $\pm 3\sigma$ is the limit of consideration in this chapter. This is partly due to the 3-sigma rule which states nearly all values lie within 3σ. But more importantly $> 3\sigma$ is counter-productive as a high tolerance will cause increased pixel-wise confusion between natural gait motion and covariate factor motion.

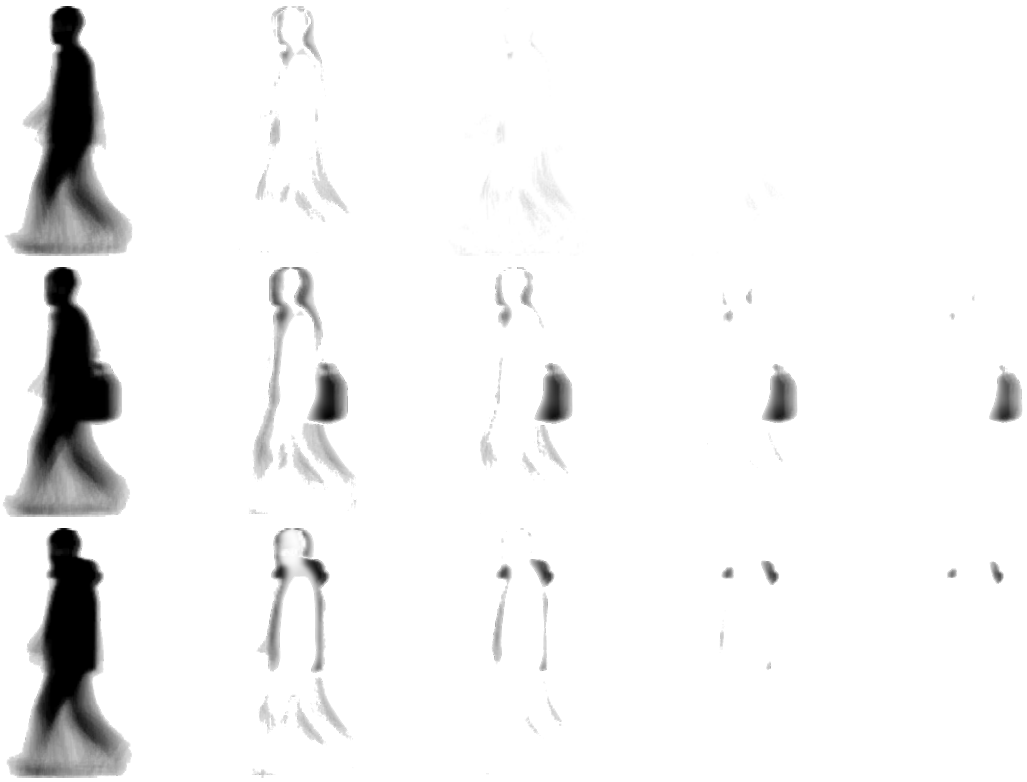


Figure 5.5: Detected covariate factors CV in the CASIA B dataset with respect to tolerance included in the "typical" GEI $tGR_{0,1,2,3}$ for normal (top), carrying a bag (middle) and clothing (bottom) sequences. Notice how a high tolerance is effective for normal sequences, however this causes increasing pixel-wise confusion between natural gait motion and covariate factor motion during covariate factor (bag and clothing) sequences

Stage 3: Covariate Factor Detection

Covariate factor detection is achieved by

$$CV(x, y) = |GR(x, y) - tGR_{0,1,2,3}(x, y)| \tag{5.8}$$

where CV are detected covariate factors, x and y are the 2D spatial image coordinates, GR is a test GR and $tGR_{0,1,2,3}$ is a "typical" GR with the relevant tolerance $\{_{0,1,2,3}\}$ applied. A visual representation of how tolerance affects covariate factor detection is seen in Figure 5.5 for the GEI in the CASIA B dataset. A covariate factor free (normal) GEI requires a high σ to incorporate the natural inter-class and intra-class variation in human gait. However, a covariate factor GEI such as carrying a bag requires a low σ. Notice how the detected bag becomes successively smaller as σ increases where the bag extracted by $\pm 3\sigma$ accounts for the static features only (i.e. it does not account for the corresponding covariate factor motion). This occurs as a high σ increases the pixel-wise confusion between natural gait motion and covariate factor motion. Conversely, when σ is low Figure 5.5 shows the GEI contains a considerably quantity of covariate factors. This is somewhat true as the bag causes leaning which therefore affects the natural appearance and motion of gait. However this demonstrates the "typical" GEI sensitivity to the natural inter-class and intra-class variance incorporated. These observations are true for all covariate factor sequences, GRs and datasets. In addition, Figure 5.5 further emphasizes that exceeding $\pm 3\sigma$ does not contribute to effective covariate factor detection.

5.3 Covariate Factor Removal

Detected covariate factors must be removed as their covariate factor free equivalent is unknown. Therefore covariate factor removal requires a three stage process: *1*) apply a threshold to the detected covariate factors in order to satisfy the trade-off for incorporating the natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion, *2*) remove the covariate factors and finally *3*) reclaim any discriminative limb-based dynamic features which have been removed by preceding stages.

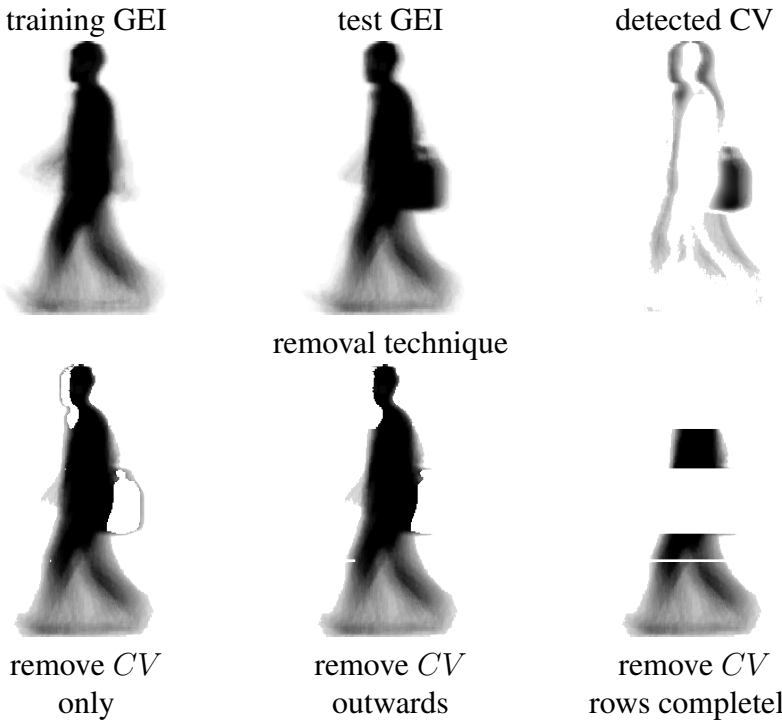


Figure 5.6: Using the GEI as an example, the detected covariate factor areas CV are removed using three removal techniques: remove CV only, remove CV outwards from the centreline of the body and remove CV rows completely; notice the similarity between the remaining pixel intensity values and the (covariate factor free) training GEI

Stage 1: Covariate Factor Threshold

As the GRs contain normalised pixel intensity values (i.e. $0 \leq GR(x,y) \leq 1$), the detected covariate factor pixel intensity values will vary based on the covariate factor present; note that these values should not be re-normalised as they are indicative of person identity. Figure 5.5 shows that *a*) a high σ is required to incorporate the natural inter-class and intra-class variance in human gait and *b*) a low σ is required for effective covariate factor detection; this occurs for the GEI, GVI, SEIM and SVIM. Therefore a threshold is required to satisfy the trade-off between incorporating the natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion. A broad range of threshold values $\{T_h = 0.1 \text{ to } 1 \text{ in steps of } 0.1\}$ are applied to the detected covariate factors to determine their contribution to effective covariate factor removal; these threshold T_h values are chosen given $0 \leq CV(x,y) \leq 1$.

Result: Remove covariate factors only

```
for every pixel value do
    if pixel value > threshold then
        set pixel value to zero;
    end
end
```

Result: Remove covariate factors outwards from the centreline of the body

```
for every row do
    calculate midpoint of the GR
    for RHS (centreline ⟶ RHS) do
        for every pixel value do
            if pixel value > threshold then
                set all pixel values in the row to zero;
            end
        end
    end
    repeat process for LHS
end
```

Result: Remove covariate factor rows completely

```
for every row do
    for every pixel value do
        if pixel value > threshold then
            set all pixel values in the row to zero;
        end
    end
end
```

Algorithm 1: Pseudocode for each covariate factor removal technique

Stage 2: Covariate Factor Removal Techniques

Robust gait recognition relies on effective covariate factor removal. Therefore three covariate factor removal techniques (pseudocode provided in Algorithm 1) are used. These are demonstrated in Figure 5.6 for the GEI in the CASIA B dataset.

Removing covariate factors only. This technique removes the detected covariate factors from a test GR. This may be ineffective if the detection stage cannot pixel-wise differentiate between natural gait motion and covariate factor motion. The effect of this limitation is seen in Figure 5.6 where an outline (covariate factor motion pixel intensity values) surrounds the removed covariate factors.

Removing covariate factors outwards from the centreline of the body. This technique is designed to resolve the limitation of removing covariate factors only. By removing detected covariate factors outwards from the centreline of the body, Figure 5.6 shows the removal of covariate factor motion which has been confused with natural gait motion.

Removing covariate factor rows completely. This technique is designed to remove covariate factors which encroach the silhouette figure. This is achieved by completely removing rows of a GR where covariate factors exist. As seen in Figure 5.6, this technique can remove a significant quantity of the GR which can risk removing discriminative features indicative of person identity.

Stage 3: Reclaiming Discriminative Leg Features

Dynamic features must be retained as they are discriminative and less sensitive to covariate factors. Therefore by considering the rows of the GR image from bottom to top, all rows are reclaimed (despite being removed by Stages 1 and 2) until a row containing a static feature (i.e. $GR(x, y) = 1$) is found; this process is similar to Bashir et al. (2008a).

5.4 Experimental Procedure

5.4.1 Dataset

The CFDR module is validated in the CASIA B and TUM GAID datasets which are explained in Chapter 2.4.

5.4.2 Validation Representations

The GEI is constructed using parameters defined in Chapter 3, while the GVI, SEIM and SVIM are constructed using parameters defined in Chapter 4. The SEIM and SVIM are derived from the screened Poisson distance function with the following smoothing parameter t: CASIA B dataset $t = 30$ and $t = 70$ respectively, TUM GAID dataset $t = 90$ and $t = 5$ respectively.

5.4.3 Dimensionality Reduction and Classification

The GEI, GVI, SEIM and SVIM use standard dimensionality reduction and classification procedures [Han and Bhanu (2006)] for single compact 2D gait representations. These procedures are effective given the small number of training sequences in gait recognition datasets. The GEI, GVI, SEIM and SVIM (standard dataset image sizes - CASIA B: 240×240, TUM GAID: 128×178) describe gait when reshaped into a 1D feature vector (CASIA B: $57600D$, TUM GAID: $22784D$). The dimensionality is reduced using Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) (code adapted from [van der Maaten (2007)]). The combination of PCA and LDA satisfies the best data representation with respect to covariance and class separation respectively (CASIA B: $123D$, TUM GAID: $154D$ account for $\approx 97\%$ variance). Nearest Neighbour (NN) classification is used with the Euclidean and Cosine distance metrics which are standards set by the CASIA B and TUM GAID datasets respectively. The CFDR module results are compared to the ground truth in a confusion matrix. The average of the diagonal in the confusion matrix yields the correct classification.

Result: Recognition procedure
for *every test GR* **do**
    detect covariate factors;
    remove covariate factors from test GR and training GRs;
    dimensionality reduction and classification;
end

Algorithm 2: Pseudocode for the recognition procedure

5.4.4 Recognition Procedure

The process of classifying each test GR is shown in Algorithm 2. Each test GR is considered individually where the covariate factors are removed from the test GR and training GRs. This approach ensures dimensionality reduction and classification is performed on covariate factor free GRs. More importantly, this process ensures a fair comparison as comparing a training GR and a test GR with covariate factors removed causes unnecessary misclassification during Nearest Neighbour classification.

5.5 Results and Discussion

The CFDR module is applied to the GEI, GVI, SEIM and SVIM and validated in the CASIA B and TUM GAID dataset. Given these results, the performance is discussed based on *1*) covariate factor effects, *2*) "typical" GR tolerance, *3*) covariate factor threshold and *4*) covariate factor removal technique.

The effects of "typical" GR tolerance and covariate factor threshold are shown in separate figures for each GR and dataset. CASIA B dataset results are presented in Figure 5.7, Figure 5.8, Figure 5.9 and Figure 5.10 for the GEI, GVI, SEIM and SVIM respectively. TUM GAID dataset results are presented in Figure 5.11, Figure 5.12, Figure 5.13 and Figure 5.14 for the GEI, GVI, SEIM and SVIM respectively. The covariate factor removal results are seen in Table 5.2 for the combination of "typical" GR tolerance and covariate factor threshold achieving the highest performance across all covariate factor sequences in the dataset. The table and figures of results show the performance before (baseline) the CFDR module is applied.

CASIA B dataset (124 test persons)

GR	Removal Technique (%)	nm	bg	cl	average
GEI	Baseline	**100.0**	53.2	22.2	58.5
	CV only	99.2	75.4	64.1	79.6
	CV outwards	99.2	76.6	65.7	80.5
	CV rows completely	98.4	**77.4**	**93.1**	**89.7**
GVI	Baseline	**97.2**	77.8	50.4	75.1
	CV only	**97.2**	78.6	50.8	75.5
	CV outwards	**97.2**	78.2	50.8	75.4
	CV rows completely	95.6	**85.9**	**71.4**	**84.3**
SEIM	Baseline	98.0	93.1	69.8	87.0
	CV only	98.0	93.1	69.8	87.0
	CV outwards	98.0	93.1	69.8	87.0
	CV rows completely	98.0	93.1	69.8	87.0
SVIM	Baseline	**98.4**	92.7	71.8	87.6
	CV only	98.0	96.4	72.6	89.0
	CV outwards	98.0	**96.8**	73.0	**89.2**
	CV rows completely	97.2	94.8	**73.8**	88.6

TUM GAID datasets (155 test persons)

GR	Removal Technique (%)	N	B	S	TN	TS	TB	average
GEI	Baseline	**99.7**	19.0	**96.5**	34.4	0.0	43.8	67.4
	CV only	98.1	53.9	88.1	**43.8**	**28.1**	37.5	75.9
	CV outwards	99.0	42.3	92.3	40.6	15.6	43.8	73.7
	CV rows completely	98.7	**58.1**	87.4	37.5	21.9	**46.9**	**77.1**
GVI	Baseline	**99.0**	47.7	94.5	62.5	15.6	**62.5**	77.3
	CV only	98.1	64.2	94.2	**65.6**	28.1	**62.5**	82.4
	CV outwards	98.4	**68.1**	**95.8**	59.4	25.0	50.0	83.4
	CV rows completely	98.7	**68.1**	93.9	62.5	**34.4**	59.4	**83.6**
SEIM	Baseline	98.7	18.4	**96.1**	31.3	0.0	31.3	66.4
	CV only	98.7	45.2	91.6	**37.5**	**25.0**	34.4	74.2
	CV outwards	**99.0**	**47.1**	92.9	**37.5**	12.5	25.0	**74.6**
	CV rows completely	98.7	45.5	92.3	**37.5**	**25.0**	**37.5**	**74.6**
SVIM	Baseline	**98.4**	64.2	**91.6**	65.6	31.3	50.0	81.4
	CV only	98.1	73.2	89.4	65.6	40.6	**56.3**	83.6
	CV outwards	**98.4**	70.3	**91.6**	71.9	34.4	53.1	83.6
	CV rows completely	**98.4**	**74.8**	89.7	**68.8**	**43.8**	43.8	**84.3**

Table 5.2: Covariate factor removal results in the CASIA B and TUM GAID datasets for each gait presentation (GR) with the highest average performing combination of "typical" GR tolerance and covariate factor threshold T_h; the baseline indicates pre-CFDR module results

5.5.1 Covariate Factor Effect on Performance

The following observations regarding the CFDR module in Table 5.2 occur regardless of gait representation (GEI, GVI, SEIM, SVIM) and dataset. Covariate factor free (normal) test GRs (CASIA B: nm; TUM GAID: N, S) achieve a high performance given the visual similarities to training GRs. The shoe sequences in the TUM GAID dataset achieve a high performance as the clean room shoe covers have a negligible impact on the natural appearance and motion of gait. The shoe sequences are aimed towards acoustic gait recognition (identifying a person based on the sounds made while walking) given the acoustic impact of the clean room shoe covers rustling. However alternative shoe types, such as flip flops [Bouchrika and Nixon (2008)] and heels, can cause increased misclassification given the impact on the natural appearance and motion of gait.

Covariate factor sequences (CASIA B: bg, cl; TUM GAID: B, TN, TB, TS) have a significant impact on the natural appearance and motion of gait which causes decreased performance. Table 5.2 shows that the TUM GAID time-based sequences (TN, TB, TS) achieve a lower performance as these are complex coupled covariate factor sequences. Notice how the performance is near half that achieved during single covariate factor sequences (N, B, S).

5.5.2 "Typical" GR Tolerance

Tolerance in the "typical" GR is used to incorporate the natural inter-class and intra-class variance in human gait. However as seen in Figure 5.5, incorporating tolerance causes pixel-wise confusion between natural gait motion and covariate factor. The CASIA B dataset "typical" GR tolerance results are seen in Figure 5.7, Figure 5.8, Figure 5.9 and Figure 5.10. The TUM GAID dataset "typical" GR tolerance results are seen in Figure 5.11, Figure 5.12, Figure 5.13 and Figure 5.14

The GEI is the least naturally robust GR seen in Table 5.2. Covariate factor free GEIs (CASIA B: nm, TUM GAID: N, S) achieve a higher performance with a high tolerance (tGR_3) to incorporate the natural inter-class and intra-class variance in human gait. Conversely, covariate factor GEIs (CASIA B: bg, cl, TUM GAID: B, TN, TB, TS) achieve a higher performance with no tolerance (tGR_0) to minimise the pixel-wise confusion be-

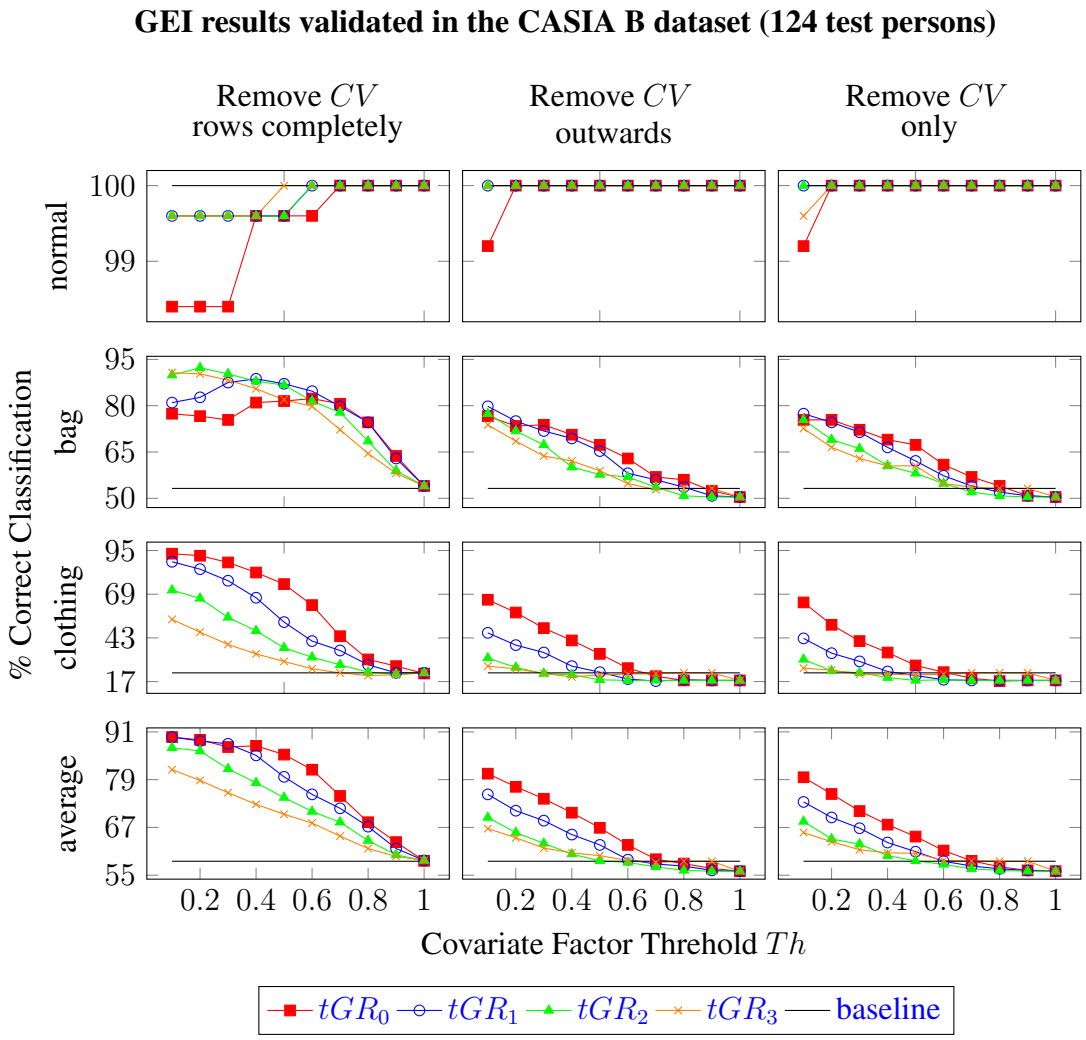


Figure 5.7: CFDR module applied to the GEI in the CASIA B dataset for each "typical" GEI tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The GEI performance is increased with tGR_0 and $T_h = 0.1$

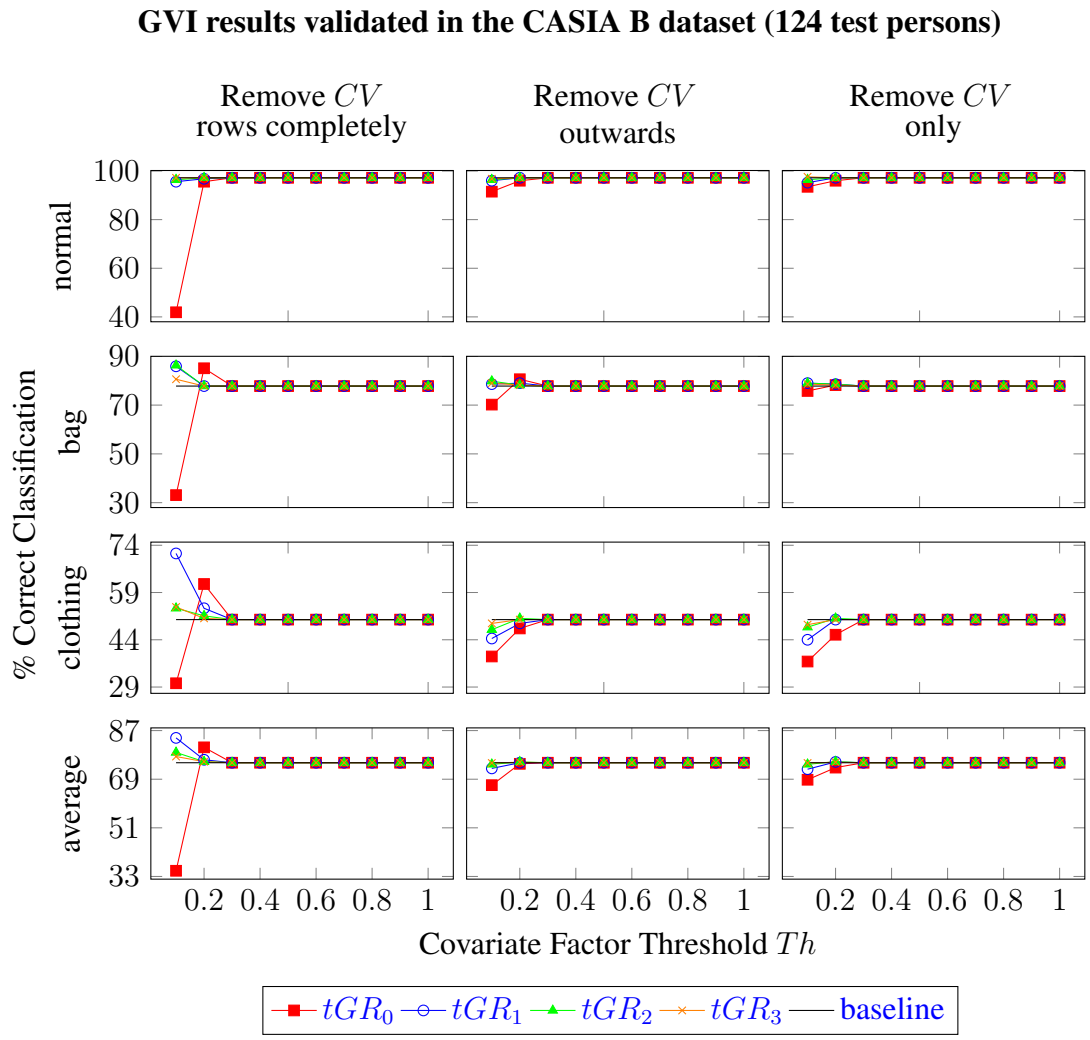


Figure 5.8: CFDR module applied to the GVI in the CASIA B dataset for each "typical" GVI tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The GVI performance is increased with tGR_1 and $T_h = 0.1$

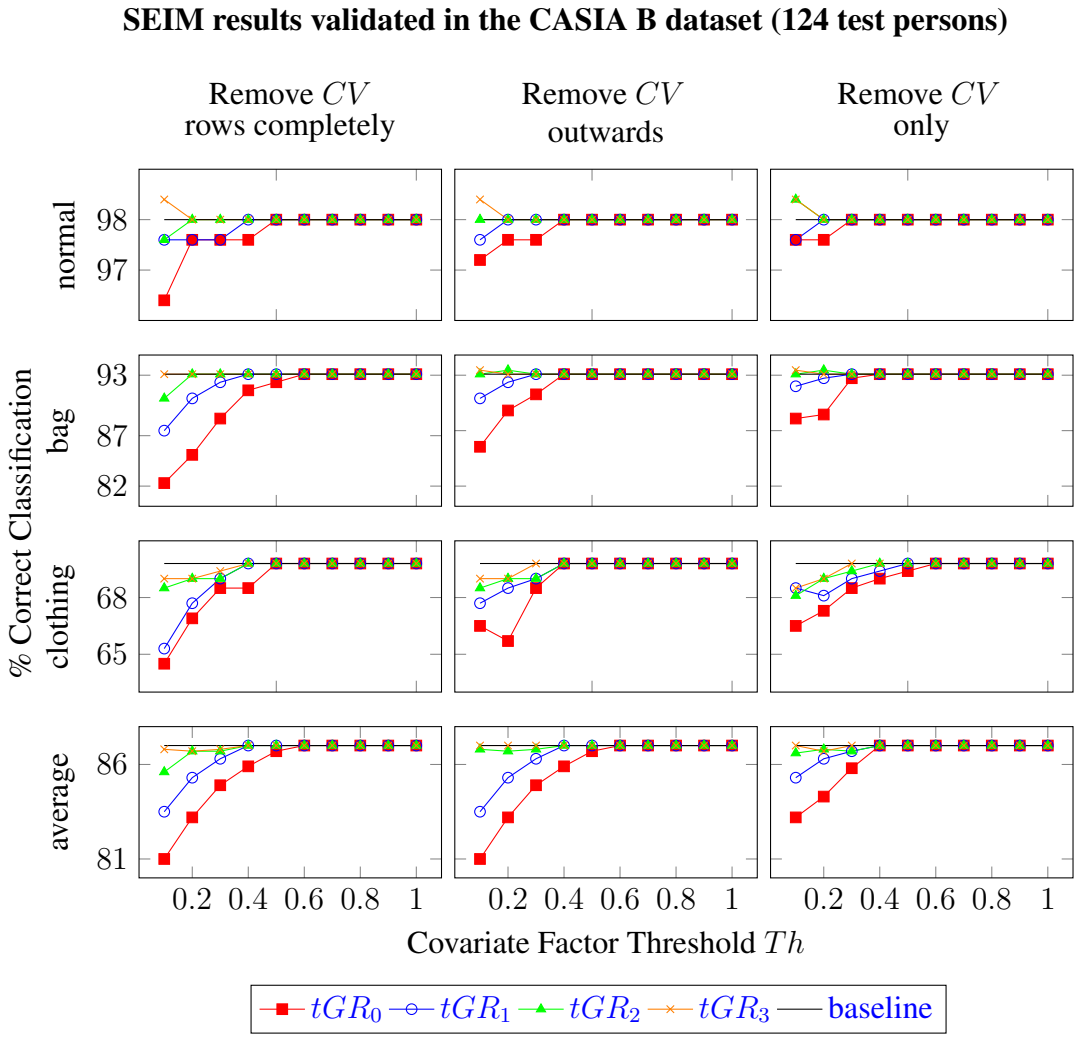


Figure 5.9: CFDR module applied to the SEIM in the CASIA B dataset for each "typical" SEIM tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The SEIM performance is not increased with the CFDR module due to the poor quality silhouettes which reduce the similarities between training SEIMs and test SEIMs

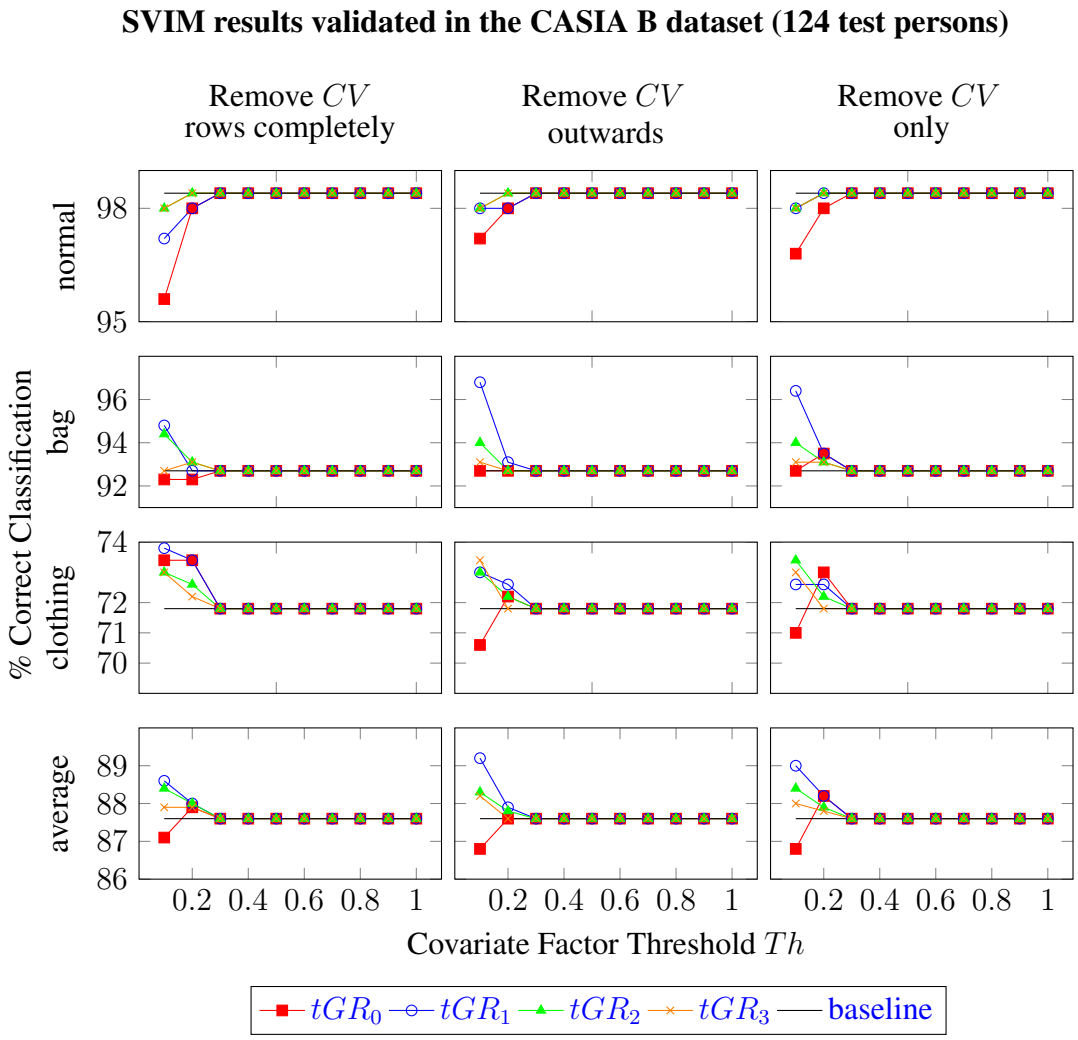


Figure 5.10: CFDR module applied to the SVIM in the CASIA B dataset for each "typical" SVIM tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The SVIM performance is increased with tGR_1 and $T_h = 0.1$

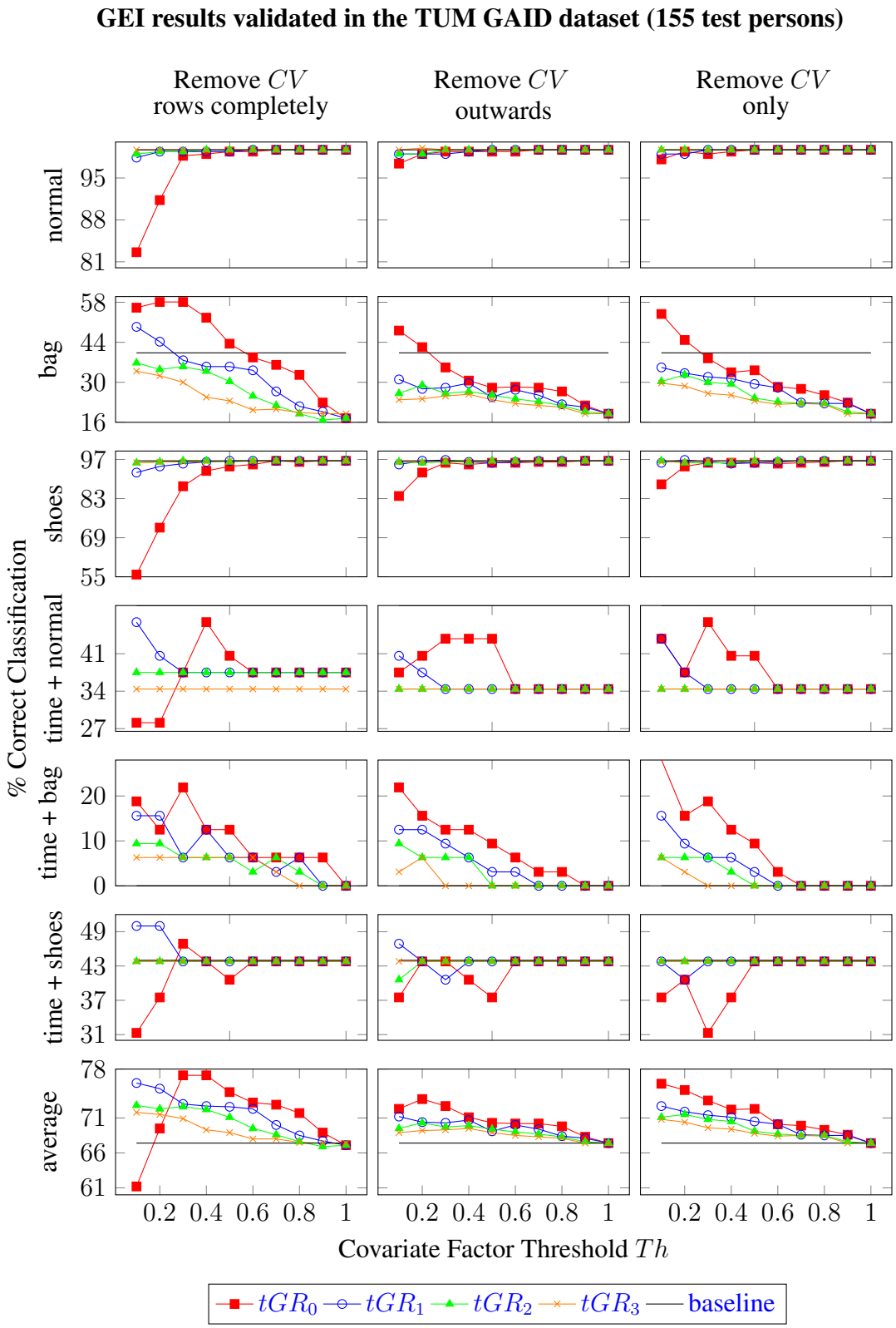


Figure 5.11: CFDR module applied to the GEI in the TUM GAID dataset for each "typical" GEI tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The GEI performance is increased with tGR_0 and $T_h = 0.3$

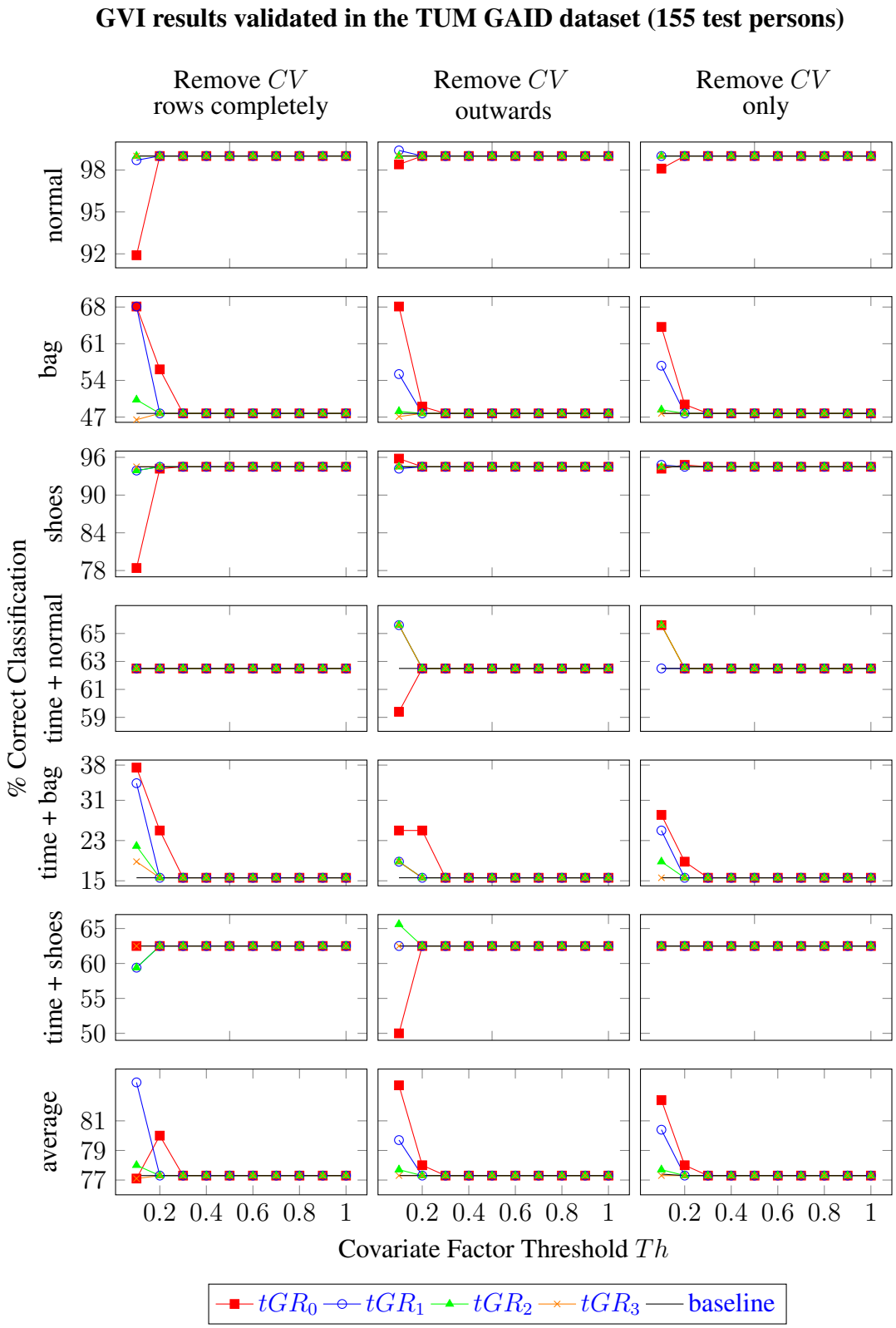


Figure 5.12: CFDR module applied to the GVI in the TUM GAID dataset for each "typical" GVI tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The GVI performance is increased with tGR_1 and $T_h = 0.1$

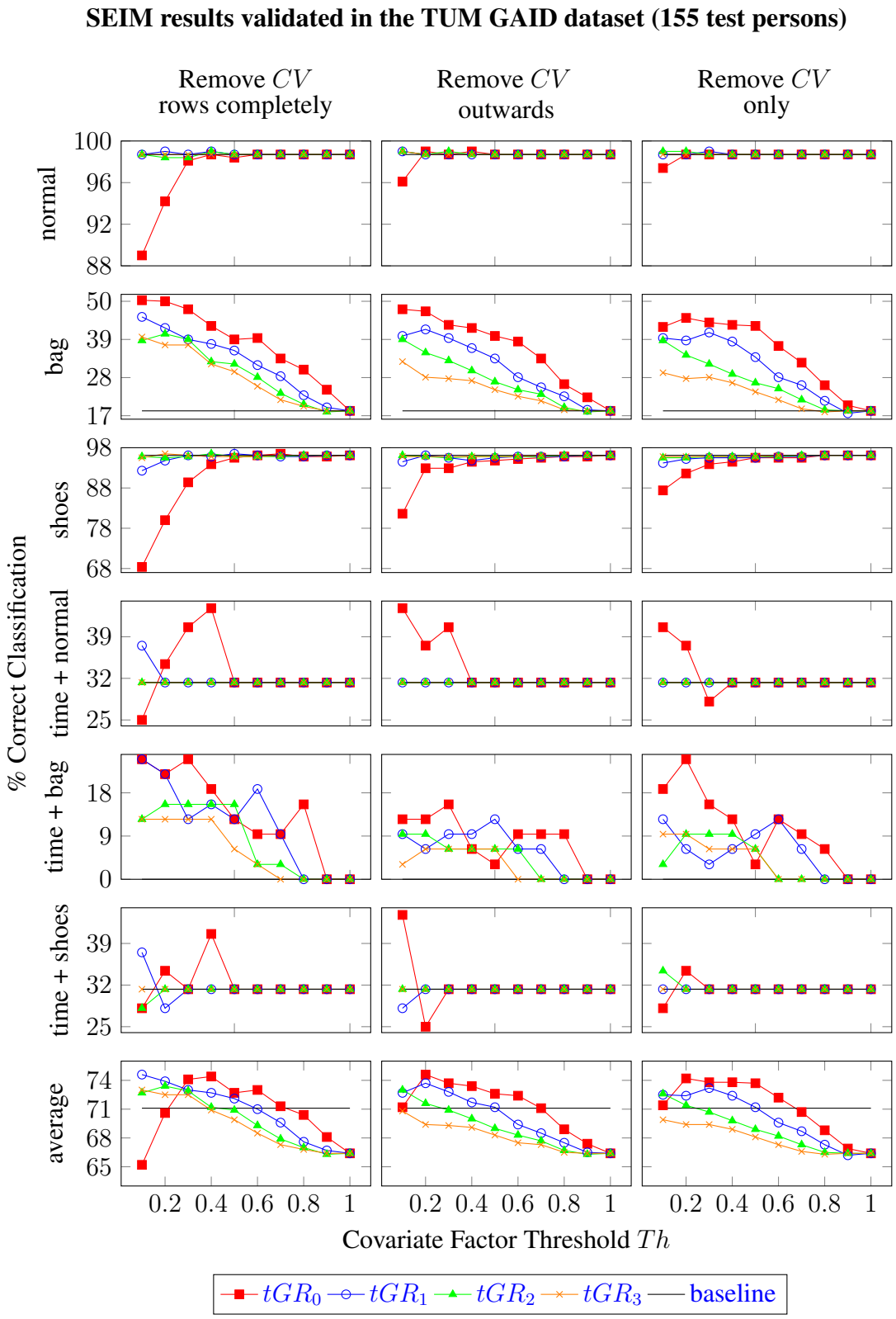


Figure 5.13: CFDR module applied to the SEIM in the TUM GAID dataset for each "typical" SEIM tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The SEIM performance is increased with tGR_1 and $T_h = 0.1$

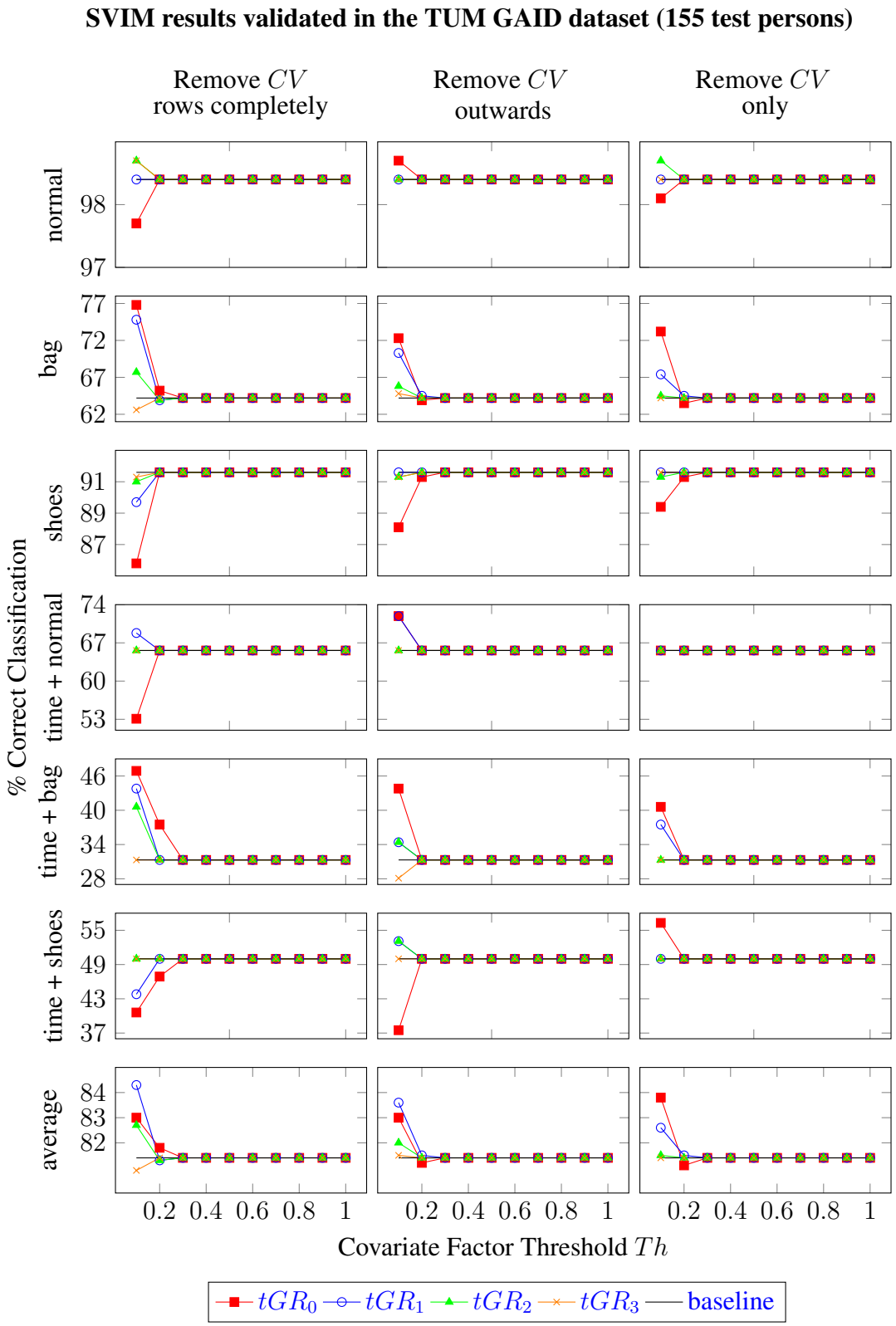


Figure 5.14: CFDR module applied to the SVIM in the TUM GAID dataset for each "typical" SVIM tolerance $tGR_{0,1,2,3}$, covariate factor threshold T_h and covariate factor removal technique; baseline performance is without applying the CFDR module. The SVIM performance is increased with tGR_1 and $T_h = 0.1$

tween natural gait motion and covariate factor motion. Therefore regardless of dataset, the GEI requires tGR_0 to achieve the highest average performance for all covariate factor sequences.

However, Table 5.2 shows that the GVI, SEIM and SVIM achieve increased robustness compared to the GEI. Covariate factor free GVIs, SEIMs and SVIMs (CASIA B: nm, TUM GAID: N, S) similarly achieve a higher performance with a high tolerance (tGR_3) to incorporate the natural inter-class and intra-class variance in human gait. The increased GVI, SEIM and SVIM robustness means that covariate factor sequences achieve a high performance with a low tolerance (tGR_1). However, notice that the SEIM in the CASIA B dataset does not achieve any performance increases; this is attributed to the sensitivity to silhouette quality which minimises the visual similarities between training SEIMs and test SEIMs. Therefore regardless of dataset, the GVI, SEIM and SVIM require tGR_1 to achieve the highest average performance across all covariate factor sequences in the CASIA B and TUM GAID dataset. By incorporating tolerance in covariate factor sequences, a satisfactory trade-off occurs between incorporating the natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion.

5.5.3 Covariate Factor Threshold

The following observations occur regardless of GR (GEI, GVI, SEIM and SVIM) and dataset (CASIA B and TUM GAID). The effect of covariate factor threshold T_h can be seen in Figure 5.7, Figure 5.8, Figure 5.9 and Figure 5.10 for the CASIA B dataset, and Figure 5.11, Figure 5.12, Figure 5.13 and Figure 5.14 for the TUM GAID dataset. The covariate factor threshold T_h is required to further satisfy the trade-off between incorporating the natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion. Similar to "typical" GR tolerance, the covariate factor threshold T_h achieving a high performance varies between covariate factor free sequences and covariate factor sequences.

Covariate factor free sequences (CASIA B: nm, TUM GAID: N, S) require a high covariate factor threshold T_h to maximise the inter-class and intra-class variance incor-

porated in the GR. Conversely, covariate factor sequences (CASIA: bg, cl, TUM GAID: B, TN, TB, TS) require a low covariate factor threshold T_h to minimise the pixel-wise confusion between natural gait motion and covariate factor motion.

To achieve the highest performance across covariate factor sequences, the GVI, SEIM and SVIM require $T_h = 0.1$ for the CASIA B and TUM GAID dataset. This is despite the different image sizes (standard dataset image sizes CASIA B: 240×240, TUM GAID: 128×178). However the GEI requires $T_h = 0.1$ and $T_h = 0.3$ for the CASIA B and TUM GAID datasets respectively. This difference may be attributed to a combination of factors e.g. *a*) silhouette quality, *b*) extracting static features and dynamic features and *c*) the pixel-wise confusion between natural gait motion and covariate factor motion. Note that these covariate factor threshold T_h values are evaluated for gait recognition and will require further evaluation for analogous applications.

5.5.4 Covariate Factor Removal Technique

The covariate factor removal stage is the final opportunity to remove covariate factors which may have previously avoided detection. Table 5.2 shows the combination of "typical" GR tolerance and covariate factor threshold T_h achieving the highest performance across each covariate factor sequence in the datasets. Figure 5.6 refers to a visualisation of each covariate factor removal technique.

Removing covariate factors only. This technique is low and middle ranking depending on the dataset. The performance is limited if the covariate factor detection stage shows pixel-wise confusion occurs between natural gait motion and covariate factor motion. This is demonstrated in Figure 5.6 where covariate factor motion pixel values surround the removed covariate factor.

Removing covariate factors outwards from the centreline of the body. While this technique visually resolves the limitation of removing covariate factors only, this is not reflected in the performance. Removing covariate factors only and removing covariate factors outwards from the centreline of the body neglect an important point. Covariate factors can lie within, and at the boundary of, the human figure silhouette. This increases

the complexity of differentiating between natural gait motion and covariate factor motion. As such, this technique is middle and low ranking depending on the dataset.

Removing covariate factor rows completely. This technique can remove a significant quantity of the GR (see Figure 5.6) and is somewhat sensitive to the natural inter-class and intra-class variance in human gait. However Table 5.2 shows this technique achieves a high performance due to the *a*) minimal pixel-wise confusion between natural gait motion and covariate factor motion and *b*) removal of covariate factors which encroach the silhouette figure.

The "remove covariate factor rows completely" technique achieves the highest performance for the majority of the GRs and datasets. The exception to this is the SVIM in the CASIA B dataset where "remove covariate factors outwards from the centreline of the body" technique is required; this may be due to the poorer silhouette quality. By using the highest performing covariate factor removal technique in Table 5.2, GR covariate factor sequence performance is increased by 15.1% with the CFDR module applied.

5.6 Comparison to State-of-the-Art

The best performing CFDR module parameters *i*) "typical" GR tolerance, *ii*) covariate factor threshold T_h and *iii*) covariate factor removal technique, for the CASIA B and TUM GAID datasets are compared to current state-of-the-art results in Table 5.3 and Table 5.4 respectively.

The CFDR module successfully sets new state-of-the-art results

CASIA B carrying a bag (bg) + 4.0%, clothing (cl) + 11.0%

TUM GAID bag (B) + 16.5%, time + normal (TN) + 4.9%, time + bag (TB) + 39.9%, time + shoes (TS) + 18.8%, weighted average + 3.6%

where on average, the CFDR module yields a 15.1% increase during covariate factor sequences.

For the CASIA B dataset, new state-of-the-art results are set for individual covariate factor sequences (bg, cl), but not for the covariate factor free sequences. This is attributed

CASIA B dataset (124 test persons)				
Gait Representation (%)	nm	bg	cl	average
CGI Wang et al. (2012)	88.1	43.7	43.0	58.2
GEI Han and Bhanu (2006)	100.0	53.2	22.2	58.5
GEnI Bashir et al. (2010)	**100.0**	78.3	44.0	74.1
MII + MDIs Bashir et al. (2009b)	97.5	83.6	48.8	76.6
SEI + GSP Huang and Boulgouris (2012)	99.0	64.0	72.0	78.3
P_{RW}GEI Yogarajah et al. (2011)	98.4	93.1	44.4	78.6
Body segmentation Li et al. (2010)	99.2	80.6	75.8	85.2
AEI Zhang et al. (2010)	98.4	91.9	72.2	87.5
SGEI + GEI Li and Chen (2013)	98.2	80.7	83.9	87.6
M_G Bashir et al. (2008a)	**100.0**	91.0	80.6	**90.5**
GEI + CFDR module	98.4	77.4	**93.1**	89.7
GVI + CFDR module	95.6	85.9	71.4	84.3
SVIM + CFDR module	98.0	**96.8**	73.0	89.2

Table 5.3: Comparison of the CFDR module results to current state-of-the-art results in the CASIA B dataset for normal (nm), carrying a bag (bg), clothing (cl). The CFDR module increases covariate factor sequences performance

TUM GAID dataset (155 test persons)

Gait Representation (%)	N	B	S	TN	TB	TS	average
depth GEI Han and Bhanu (2006)	**99.7**	17.4	**96.5**	37.5	0.0	43.8	67.1
GEV Hofmann et al. (2013)	94.2	13.9	87.7	41.0	0.0	31.0	61.4
DGHEI Hofmann et al. (2013)	99.0	40.3	96.1	50.0	0.0	44.0	74.1
GEI + CFDR module	98.7	58.1	87.4	37.5	21.9	46.9	77.1
GVI + CFDR module	98.7	68.1	93.9	62.5	34.4	**59.4**	83.6
SEIM + CFDR module	98.7	45.5	92.3	37.5	25.0	37.5	74.6
SVIM + CFDR module	98.4	**74.8**	89.7	**68.8**	**43.8**	43.8	**84.3**

Table 5.4: Comparison of the CFDR module results to current state-of-the-art results in the TUM GAID dataset for normal (N), carrying a bag (B), shoes (S), time and normal (TN), time and bag (TB), time and shoes (TS). The CFDR module significantly increases covariate factor sequences performance

to the parameter trade-off required to achieve superior covariate factor sequence performance; this incurs a minor performance drop during covariate factor free sequences. However for the TUM GAID dataset, significant state-of-the-art results are set for individual covariate factor sequences (B, TN, TB, TS) and the highest average result across covariate factors. Similar to CASIA B dataset results, covariate factor free sequences incur minor performance drops due to boosting covariate factor performance. Across datasets, the CFDR module enhances GR robustness due to satisfying the trade-off for incorporating natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion.

5.7 Conclusion

The success of the CFDR module is due to two primary reasons.

1. While covariate factors are static with respect to the body, the natural motion of gait causes covariate factors to subsequently undergo motion. This causes pixel-wise confusion between natural gait motion and covariate factor motion. Therefore, by determining the underling pixel-wise composition of covariate factors, the

CFDR module maximises covariate factor detection and removal. This ensures a favourable trade-off between incorporating the natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion. The CFDR module can also deal with covariate factors at the boundary of, and hidden within, the silhouette figure.

2. The CFDR module can remove a significant quantity of a test GR. Therefore direct classification with training GRs is not a fair comparison. By considering each test GR in turn, covariate factors are removed from both test and training GRs to ensure dimensionality reduction and classification is performed only on covariate factor free areas.

Hypothesis Revised

At the beginning of this chapter, the following hypothesis was made:

> "This chapter argues that single compact 2D gait representations can achieve superior robustness when performing dedicated covariate factor detection and removal."

Therefore during the course of this chapter, the hypothesis has been verified.

Future Directions

A further development could be made to increase the robustness of the CFDR module and also widen its application.

- There is no covariate factor threshold T_h which generalises over all covariate factors. This is natural given each covariate factor affects the natural appearance and motion of gait uniquely. This issue could be alleviated by detecting the covariate factor type and then applying the required covariate factor threshold T_h. Note that covariate factor detection would be best suited to RGB images as colour and texture cures are rejected when using silhouettes.

The most important aspect of the CFDR module is the potential application to analogous single compact 2D gait, or even action, representations due to the effectiveness, simplicity, and ability to easily fit within existing procedures.

Chapter 6

Conclusion

This chapter summarises the thesis and its contributions to knowledge. To conclude the thesis, a number of future directions and open research problems are highlighted.

6.1 Thesis Summary

This goal of this thesis is to develop a set of techniques to boost the robustness of gait recognition.

This is achieved by establishing the simplifying assumptions which have thus limited gait recognition robustness. Current research identifies covariate factors as static with respect to the body. While this is true, research neglects the fact that covariate factors subsequently undergo motion due to the natural motion of human gait. Chapter 4 and Chapter 5 specifically develop novel gait representations and descriptors which are capable of differentiating between natural gait motion and covariate factor motion. By performing validation with datasets containing high test person numbers and real world complex covariate factors, this thesis achieves significant state-of-the-art advances.

Chapter 2 indicates model-free approaches containing silhouette (or derivatives thereof) single compact 2D representations are powerful and robust gait recognition tools with reduced computational demands. Analysing gait recognition literature exposes two key limitations, namely *i*) while covariate factors are static with respect to the body, natural gait motion causes covariate factors to subsequently undergo motion, and this leads to

ineffective covariate factor removal due to a knowledge gap in how covariate factors are composed at the pixel intensity value level and *ii*) a failure to consider the degree and severity to which covariate factors uniquely affect the natural appearance and motion of gait.

Chapter 3 demonstrates how the action recognition and gait recognition performance is affected when using HOG as a "black box". The HOG gradient scheme, cell size and bin size are evaluated for person detection but are subsequently used on other applications without further analysis. For the alternative applications of action recognition and gait recognition, the following HOG parameters are evaluated *i*) eight gradient schemes with varying gradient orientation and gradient magnitude accuracy and *ii*) 100 cell size and bin size combinations. The optimal parameters for person detection are compared against optimal parameters for action recognition and gait recognition.

Chapter 4 promotes the use of fuzzy skeletons in gait recognition. To overcome boundary noise sensitivity, three Poisson-based smooth distance functions are evaluated to determine which accuracy (near or far from the boundary) properties are required for gait recognition. The fuzzy skeleton sequence is condensed into a single compact 2D gait representation, named the Skeleton Variance Image (SVIM), by computing the pixel-wise variance. The performance of the SVIM is compared against analogous fuzzy skeleton representations and silhouette representations.

Chapter 5 maximises covariate factor detection and removal in single compact 2D gait representations with the CFDR module. Covariate factor detection is achieved by determining the average pixel composition during covariate factor free sequences. The 3-sigma rule is applied to determine the trade-off between incorporating the natural inter-class and intra-class variance in human gait versus minimising the pixel-wise confusion between natural gait motion and covariate factor motion. Three covariate factor removal techniques are evaluated to determine their effectiveness in removing covariate factors which lie at, and hidden within, the boundary of the silhouette figure. The performance compar-

ison is based on GRs with and without the CFDR module applied.

Therefore, this thesis achieves the goal of enhanced gait recognition robustness using a combination of innovative techniques and tangible solutions. Throughout this thesis, an important and consistent conclusion indicates there is no single solution or parameter for robust gait recognition. This is expected given the unique manner in which covariate factors affect the natural appearance and motion of gait.

6.2 Contributions

The novel contributions of this thesis are summarised.

Gait Energy Image Described by Histograms of Oriented Gradients

Presented at the British Machine Vision Conference Student Workshop [Whytock et al. (2012)] and the International Symposium on Visual Computing [Whytock et al. (2013a)].

Hypothesis

This chapter argues that to ensure robustness, HOG parameters (gradient scheme, cell size and bin size) must undergo re-evaluation when applied to applications other than person detection.

HOG is a highly cited feature descriptor used by numerous computer vision applications. However HOG is commonly used as a "black box" meaning parameters evaluated for person detection are used for other applications without re-evaluation. For the alternative applications of action recognition and gait recognition, Wilcoxon tests indicate that cell size and bin size have a significant effect on performance compared to the gradient scheme. Overall, the combination of GEI and HOG is effective for action recognition when validated in the Weizmann Action dataset. However the GEI and HOG combination does not yield a satisfactory performance for gait recognition when validated in the CASIA B and TUM GAID dataset. The poorer performance occurs as *a*) the combination does not scale with dataset size and *b*) HOG encodes the appearance and motion of covariate factors in the GEI.

Variance-based Fuzzy Skeletal Features

Published in the Journal of Mathematical Imaging and Vision [Whytock et al. (2014)].

Hypothesis

This chapter argues that by exploiting the Poisson equation to construct a smooth distance function, fuzzy skeletons can be extracted and formed into a single compact 2D gait representation to yield a discriminative gait descriptor.

The SVIM exploits the novel combination of skeleton and single compact 2D gait representations. A smooth distance function derived from the screened Poisson equation is used to absorb boundary noise; the addition of a tunable smoothing parameter is effective for covariate factor generalisation. A covariate factor manifests itself as a bend in the fuzzy skeleton which ensures the mitigation of covariate factor motion in the SVIM. The fuzzy skeleton sequence is condensed into the SVIM by computing the pixel-wise variance; this process extracts discriminative motion features. Therefore the SVIM shows a 9.9% increase over state-of-the-art when validated in the TUM GAID dataset.

Covariate Factor Detection and Removal

Presented at the International Symposium on Visual Computing [Whytock et al. (2013c)] and the International Conference on Imaging for Crime Detection and Prevention [Whytock et al. (2013b)], and published in the Journal of Machine Vision and Applications [Whytock et al. (2015)].

Hypothesis

This chapter argues that single compact 2D gait representations can achieve superior robustness when performing dedicated covariate factor detection and removal.

The CFDR module detects and removes covariate factors in single compact 2D gait representations. By establishing the pixel intensity distribution in covariate factor free training sequences, a degree a tolerance can be included to incorporate the natural inter-class and intra-class variance in human gait. This process maximises covariate factor detection and minimises the pixel-wise confusion between natural gait motion and covariate factor

motion. Effective covariate factor removal occurs by removing complete rows where covariate factors exists. This means the CFDR module can remove covariate factors at the boundary of, and hidden within, the silhouette figure. Given the state-of-the-art SVIM from Chapter 4, the application of the CFDR module shows a further 3.6% increase when validated in the TUM GAID dataset.

This thesis establishes and exploits the current limitations of existing gait recognition research. As a consequence, this thesis demonstrates a number of techniques which enhance gait recognition robustness. As a research topic, gait recognition will continue to be an open and interesting problem due to the numerous real world applications and the infancy of gait as a biometric (compared to well established biometrics such as fingerprint).

6.3 Future Research Directions and Open Problems

A number of future directions and open problems exist for gait recognition research. These are based on enhancing robustness, establishing the limitations of gait recognition, and prompting the next generation of validation datasets.

Covariate factor detection

This thesis demonstrates there is no single solution or parameter for robust gait recognition. This is expected given covariate factors uniquely affect the natural appearance and motion of gait. The ability to initially recognise covariate factors could significantly boost robustness as the optimum solution or parameter could be applied. Note that covariate factor classification is a non trivial task given their various shapes, sizes and effects. Covariate factor recognition could be achieved by retaining the discriminative colour and texture cues. Therefore if a known human figure shape is established, any deviations could be attributed to covariate factors and thus segmented and classified.

Establishing the limits of gait recognition

Gait recognition is concerned with identifying a person by their unique walking manner. However it is important to establish whether gait recognition extends to other actions

e.g. running. This is essential as it cannot be guaranteed that a person will be captured walking. This can be achieved using action recognition to sub-divide image sequences based on the action performed. This will require a new dataset which captures persons performing multiple actions (including walking) in a similar environment.

It is essential to establish the potential vulnerabilities in gait recognition e.g. imitating the gait of another person. Considering the forensic applications of gait recognition, a criminal may imitate the gait of a person to evade recognition; however when fleeing a crime scene the "fight or flight" mode may cause the criminal to revert to their own gait. This will require a new dataset to establish whether imitating the gait of another person could cast reasonable doubt towards the identity of a person.

Enhanced datasets

It is essential for gait recognition to perform with person numbers reflecting existing biometrics, e.g. fingerprint uses a minimum of 1000 test persons. In addition, more real world covariate factors are required for enhanced validation. Example covariate factors to consider include capture *i*) when music is playing, *ii*) when a person is intoxicated (drunkenness) or pregnant, *iii*) when people walk as part of a crowd and *iv*) a greater focus on elapsed time (years instead of months) between capture. To push gait recognition into real world deployment, these factors should be incorporated in the next generation of gait recognition datasets.

Bibliography

Aggarwal, J. and M. Ryoo (2011). Human activity analysis: A review. *ACM Computing Surveys 43*(3), 16:1–16:43.

Ambrosio, L. and V. M. Tortorelli (1990). Approximation of functionals depending on jumps by elliptic functionals via γ-convergence. *Communications on Pure and Applied Mathematics XLIII*, 999–1036.

Aubert, G. and J.-F. Aujol (2012). Poisson skeleton revisited: a new mathematical perspective. *Journal of Mathematical Imaging and Vision*, 1–11.

Aubert, G. and P. Kornprobst (2002). *Mathematical Problems in Image processing*. Springer.

Bashir, K., T. Xiang, and S. Gong (2008a). Feature Selection for Gait Recognition without Subject Cooperation. In *British Machine Vision Conference*.

Bashir, K., T. Xiang, and S. Gong (2008b). Feature selection on Gait Energy Image for human identification. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, pp. 985 – 988.

Bashir, K., T. Xiang, and S. Gong (2009a). Gait recognition using Gait Entropy Image. In *Crime Detection and Prevention, International Conference on*, pp. 1 – 6.

Bashir, K., T. Xiang, and S. Gong (2009b). Gait representation using flow fields. In *British Machine Vision Conference*.

Bashir, K., T. Xiang, and S. Gong (2010). Gait recognition without subject cooperation. *Pattern Recognition Letters 31*(13), 2052–2060.

Belyaev, A. (2011). On implicit image derivatives and their applications. In *British Machine Vision Conference*, pp. 1 – 12.

Belyaev, A. (2013). Implicit image differentiation and filtering with applications to image sharpening. *SIAM Journal on Imaging Sciences 6*(1), 660679.

Bickley, W. G. (1948). Finite difference formulae for the square lattice. *Quarterly Journal of Mechanics and Applied Mathematics 1*, 35–42.

Blank, M., L. Gorelick, E. Shechtman, M. Irani, and R. Basri (2005). Actions as Space-time Shapes. In *Computer Vision, IEEE International Conference on*, Volume 2, pp. 1395 – 1402.

Blum, H. (1967). Transformation for extracting new descriptors of shape. In W. Wathen-Dunn (Ed.), *Models for the Perception of Speech and Visual Form*. MIT Press.

Bobick, A. and J. Davis (2001). The recognition of human movement using temporal templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 23*(3), 257 – 267.

Bouchrika, I., M. Goffredo, J. Carter, and M. Nixon (2011). On using gait in forensic biometrics. *Journal of Forensic Sciences 56*(4), 882 – 889.

Bouchrika, I. and M. Nixon (2008). Exploratory factor analysis of gait recognition. In *Automatic Face and Gesture Recognition, IEEE International Conference on*.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 8*, 679–698.

Cao, L., M. Dikmen, Y. Fu, and T. Huang (2008). Gender recognition from body. In *Multimedia, Proceedings of the ACM International Conference on*, pp. 725 – 728.

Chai, Y., J. Ren, W. Han, and H. Li (2011). Human gait recognition: Approaches, datasets and challenges. In *Imaging for Crime Detection and Prevention, International Conference on*, pp. 1–6.

Chaquet, J., E. Carmona, and A. Fernndez-Caballero (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding 117*(6), 633 – 659.

Chaudhry, R., A. Ravichandran, G. Hager, and R. Vidal (2009a). Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1932 – 1939.

Chaudhry, R., A. Ravichandran, G. Hager, and R. Vidal (2009b). Histograms of Oriented Optical Flow and Binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1932 – 1939.

Chen, C., J. Liang, H. Zhao, H. Hu, and J. Tian (2009). Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters 30*(11), 977 – 984.

Chen, C.-C. and J. Aggarwal (2009). Recognizing human action from a far field of view. *Motion and Video Computing, IEEE Workshop on*.

Chen, H.-S., H.-T. Chen, Y.-W. Chen, and S.-Y. Lee (2006). Human action recognition using Star Skeleton. In *Video Surveillance and Sensor Networks, Proceedings of the ACM International Workshop on*, pp. 171 – 178.

Collins, R., R. Gross, and J. Shi (2002). Silhouette-based human identification from body shape and gait. In *Automatic Face and Gesture Recognition, IEEE International Conference on*.

Cortes, C. and V. Vapnik (1995). Support-vector networks. In *Machine Learning*, pp. 273 – 297.

Crane, K., C. Weischedel, and M. Wardetzky (2013). Geodesics in heat: A new approach to computing distance based on heat flow. *Graphics, ACM Transactions on 32*, 152:1–152:11.

Cremers, D. (2013). Shape priors for image segmentation. In *Shape Perception in Human and Computer Vision*, Advances in Computer Vision and Pattern Recognition, pp. 103 – 117. Springer London.

Cunado, D., M. Nixon, and J. Carter (1997). Using gait as a biometric, via phase-weighted magnitude spectra. In *Audio- and Video-Based Biometric Person Authentication, Proceedings of International Conference on*, pp. 95–102. Springer Verlag.

Cuntoor, N., A. Kale, and R. Chellappa (2003). Combining multiple evidences for gait recognition. In *Acoustics, Speech, and Signal Processing, Proceedings of the IEEE International Conference on*, Volume 3, pp. III–33–6 vol.3.

Cutting, J. and L. Kozlowski (1977). Recognising friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society 9*(5), 353–356.

Dalal, N. and B. Triggs (2005). Histograms of Oriented Gradients for human detection. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, Volume 1, pp. 886 – 893.

Dalal, N., B. Triggs, and C. Schmid (2006). Human detection using oriented histograms of flow and appearance. In *European conference on Computer Vision Part II, Proceedings of the*, pp. 428 – 441.

Dempster, W. and G. Gaughran (1967). Properties of body segments based on size and weight. *American Journal of Anatomy 120*(1), 33 – 54.

DiMaggio, J. and W. Vernon (2011). *Forensic Podiatry*. Humana Press.

Direkoglu, C., R. Dahyot, and M. Manzke (2012). On using anisotropic diffusion for skeleton extraction. *International Journal of Computer Vision 100*, 170–189.

Dollar, P., V. Rabaud, G. Cottrell, and S. Belongie (2005). Behavior recognition via sparse spatio-temporal features. In *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Joint IEEE International Workshop on*, pp. 65 – 72.

Drillis, R. and R. Contini (1966). Body segment parameters. Office of Vocational Rehabilitation, Department of Health, Education and Welfare, New York. Report No. 1163.03.

Efros, A., A. Berg, G. Mori, and J. Malik (2003). Recognizing action at a distance. In *Computer Vision, Proceedings of the IEEE International Conference on*, Volume 2, pp. 726–733.

Evans, L. C. (1998). *Partial Differenetial Equations*. American Mathematical Society.

Fathi, A. and G. Mori (2008). Action recognition by learning mid-level motion features. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1 – 8.

Gafurov, D. (2007). A survey of biometric gait recognition: Approaches, security and challenges. In *Nik Conference*.

Gavrila, D. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding 73*(1), 82 – 98.

Gkalelis, N., H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas (2009). The i3DPost multi-view and 3D human action/interaction database. In *Visual Media Production, Conference for*, pp. 159–168.

Gorelick, L., M. Blank, E. Shechtman, M. Irani, and R. Basri (2007a). Actions as Space-time Shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 29*(12), 2247 – 2253.

Gorelick, L., M. Blank, E. Shechtman, M. Irani, and R. Basri (2007b). Actions as space-time shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*(12), 2247–2253.

Gorelick, L., M. Galun, E. Sharon, R. Basri, and A. Brandt (2004). Shape representation and classification using the poisson equation. In *Computer Vision and Pattern Recognition, Proceedings of the*, pp. 61 – 67.

Gorelick, L., M. Galun, E. Sharon, R. Basri, and A. Brandt (2006). Shape representation and classification using the Poisson equation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*(12), 1991–2005.

Gross, R. and J. Shi (2001). The CMU motion of body (mobo) database. Technical report, Carnegie Mellon University.

Gurumoorthy, K. S. and A. Rangarajan (2009). A Schrödinger equation for the fast computation of approximate Euclidean distance functions. In *Scale Space and Variational Methods in Computer Vision. LNCS*, Volume 5567, pp. 100–111. Springer.

Han, J. and B. Bhanu (2006). Individual Recognition using Gait Energy Image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 28*(2), 316 – 322.

Hofmann, M., S. Bachmann, and G. Rigoll (2012). 2.5D gait biometrics using the Depth Gradient Histogram Energy Image. In *Biometrics: Theory, Applications and Systems, IEEE International Conference on*, pp. 399 – 403.

Hofmann, M., J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll (2013). The TUM Gait from Audio, Image and Depth (GAID) Database: Multimodal Recognition of Subjects and Traits. *Journal of Visual Communication and Image Representation, Special Issue on Visual Understanding and Applications with RGB-D Cameras*.

Hofmann, M., S. Schmidt, A. Rajagopalan, and G. Rigoll (2012). Combined face and gait recognition using alpha matte preprocessing. In *Biometrics, IAPR International Conference on*, pp. 390–395.

Hofmann, M., S. Sural, and G. Rigoll (2011). Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In *Computer Graphics, Visualization and Computer Vision, International Conferences on*.

Hsu, C.-H. and C.-J. Lin (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on 13*(2), 415 – 425.

Hu, W., T. Tan, L. Wang, and S. Maybank (2004). A survey on visual surveillance of

object motion and behaviors. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 34*(3), 334–352.

Huang, X. and N. Boulgouris (2012). Gait Recognition with Shifted Energy Image and Structural Feature Extraction. *Image Processing, IEEE Transactions on 21*(4), 2256 – 2268.

Jähne, B., H. Scharr, and S. Körkel (1999). Principles of filter design. In *Handbook of Computer Vision and Applications*, Volume 2, Signal Processing and Applications, pp. 125 – 151. Academic Press.

Jain, A., A. Ross, and S. Prabhakar (2004). An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on 14*(1), 4–20.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*(2), 201 – 211.

Johnson, A. and A. Bobick (2001). A multi-view method for gait recognition using static body parameters. In *Audio- and Video-Based Biometric Person Authentication, Proceedings of the International Conference on*, pp. 301–311. Springer-Verlag.

Kale, A., A. Chowdhury, and R. Chellappa (2003). Towards a view invariant gait recognition algorithm. In *Advanced Video and Signal Based Surveillance, Proceedings of the IEEE Conference on*, pp. 143–150.

Kale, A., N. Cuntoor, and R. Chellappa (2002). A framework for activity-specific human identification. In *Acoustics, Speech, and Signal, Processing of the IEEE International Conference on*, Volume 4, pp. IV–3660–IV–3663.

Kellokumpu, V., G. Zhao, S. Li, and M. Pietikäinen (2009). Dynamic texture based gait recognition. In *Advances in Biometrics*, Volume 5558 of *Lecture Notes in Computer Science*, pp. 1000–1009. Springer Berlin Heidelberg.

Kellokumpu, V., G. Zhao, and M. Pietikäinen (2008). Human activity recognition using a dynamic texture based method. In *British Machine Vision Conference*.

Kläser, A., M. Marszalek, and C. Schmid (2008). A spatio-temporal descriptor based on 3D-gradients. In *British Machine Vision Conference, Proceedings of the*.

Kozlowski, L. and J. Cutting (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics 21*(6), 575–580.

Kuehne, H., H. Jhuang, E. Garrote, T. Poggio, and T. Serre (2011). HMDB: a large video database for human motion recognition. In *Computer Vision, Proceedings of the International Conference on*.

Lam, T., C. K.H., and J. Liu (2011). Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition 44*(4), 973 – 987.

Lam, T. and R. Lee (2005). A new representation for human gait recognition: Motion silhouettes image (msi). In *Advances in Biometrics*, Volume 3832 of *Lecture Notes in Computer Science*, pp. 612–618. Springer Berlin Heidelberg.

Laptev, I. and T. Lindeberg (2003). Space-time interest points. In *Computer Vision, Proceedings of the IEEE International Conference on*, Volume 1, pp. 432 – 439.

Laptev, I., M. Marszalek, C. Schmid, and B. Rozenfeld (2008). Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1 – 8.

Laptev, I. and G. Mori (2010). Statistical and structural recognition of human actions. In *European Conference on Computer Vision Tutorial*.

Lee, C.-C., C.-H. Chuang, J.-W. Hsieh, M.-X. Wu, and K.-C. Fan (2011). Frame difference history image for gait recognition. In *Machine Learning and Cybernetics, 2011 International Conference on*, Volume 4, pp. 1785–1788.

Lee, L. and W. Grimson (2002). Gait analysis for recognition and classification. In *Automatic Face and Gesture Recognition, Proceedings of the IEEE International Conference on*, pp. 148–155.

Lele, S. (1992a). Compact finite difference schemes with spectral-like resolution. *Journal of Computational Physics 103*(1), 16 – 42.

Lele, S. K. (1992b). Compact finite difference schemes with spectral-like resolution. *Journal of Computational Physics 103*, 16 – 42.

Leung, M. and Y.-H. Yang (1995). First sight: A human body outline labeling system. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 17*(4), 359–377.

Li, N., Y. Xu, and X. Yang (2010). Part-based human gait identification under clothing and carrying condition variations. In *Machine Learning and Cybernetics, International Conference on*, Volume 1, pp. 268 – 273.

Li, S. (Ed.) (2009). *Encyclopedia of Biometrics*. Springer US.

Li, X. and Y. Chen (2013). Gait Recognition Based on Structural Gait Energy Image. *Journal of Computational Information Systems 9*(1), 121 – 126.

Li, X., S. Maybank, S. Y., D. Tao, and D. Xu (2008). Gait components and their application to gender recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 38*(2), 145–155.

Lin, H.-W., M.-C. Hu, and J.-L. Wu (2012). Gait-based action recognition via accelerated minimum incremental coding length classifier. In *Advances in Multimedia Modeling*, Volume 7131 of *Lecture Notes in Computer Science*, pp. 266 – 276.

Little, J. and J. Boyd (1998). Recognizing people by their gait: The shape of motion. *Videre: Journal of Computer Vision Research 1*(2), 1–32.

Liu, H., R. Feris, and M.-T. Sun (2011). Benchmarking datasets for human activity recognition. In *Visual Analysis of Humans*, pp. 411 – 427.

Liu, J., J. Luo, and M. Shah (2009). Recognizing realistic actions from videos“in the wild”. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 1996–2003.

Liu, J. and N. Zheng (2007). Gait history image: A novel temporal template for gait recognition. In *Multimedia and Expo, IEEE International Conference on*, pp. 663–666.

Liu, Y., J. Zhang, C. Wang, and L. Wang (2012). Multiple HOG templates for gait recognition. In *Pattern Recognition, International Conference on*, pp. 2930–2933.

Lu, J. and Y.-P. Tan (2010). Gait-based human age estimation. *Information Forensics and Security, IEEE Transactions on 5*(4), 761–770.

Ma, Q., S. Wang, D. Nie, and J. Qiu (2007). Recognizing humans based on gait moment image. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, ACIS International Conference on*, Volume 2, pp. 606–610.

Makihara, Y., H. Mannami, A. Tsuji, M. Hossain, K. Sugiura, A. Mori, and Y. Yagi (2012). The ou-isir gait database comprising the treadmill dataset. *Computer Vision and Applications, IPSJ Transactions on 4*, 5362.

Makihara, Y., M. Okumura, H. Iwama, and Y. Yagi (2011). Gait-based age estimation using a whole-generation gait database. In *Biometrics, International Joint Conference on*, pp. 1–6.

Marszalek, M., I. Laptev, and C. Schmid (2009). Actions in context. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 2929 – 2936.

Martin-Félez, R., R. Mollineda, and J. Sánchez (2010). Towards a more realistic appearance-based gait representation for gender recognition. In *Pattern Recognition, International Conference on*, pp. 3810–3813.

Martín-Félez, R. and T. Xiang (2012). Gait Recognition by Ranking. In *Computer Vision ECCV*, Volume 7572 of *Lecture Notes in Computer Science*, pp. 328 – 341.

Matovski, D., M. Nixon, and J. Carter (2013). *Encyclopedia of Computer Vision*, Chapter Gait Recognition. Springer Science+Business Media. (In Press).

Matovski, D., M. Nixon, S. Mahmoodi, and J. Carter (2012). The effect of time on gait recognition performance. *Information Forensics and Security, IEEE Transactions on 7*(2), 543–552.

Messing, R., C. Pal, and H. Kautz (2009). Activity recognition using the velocity histories of tracked keypoints. In *Computer Vision, IEEE International Conference on*, pp. 104–111.

Moeslund, T., A. Hilton, and V. Krüger (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding 104*(23), 90 – 126.

Mumford, D. and J. Shah (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics 42*(5), 577–685.

Murray, M., A. Drought, and R. Kory (1964). Walking Patterns of Normal Men. *The Journal of Bone and Joint Surgery 46*(2), 335–360.

Niebles, J., C.-W. Chen, and L. Fei-Fei (2010). Modeling temporal structure of decomposable motion segments for activity classification. Volume 6312 of *Lecture Notes in Computer Science*, pp. 392–405. Springer Berlin Heidelberg.

Nixon, J., J. Carter, and M. Grant (2001). Experimental plan for automatic gait recognition. Technical report, University of Southampton.

Nixon, M., I. Bouchrika, B. Arbab-Zavar, and J. Carter (2010). On use of biometrics in forensics: gait and ear. In *European Signal Processing Conference*.

Niyogi, S. and E. Adelson (1994). Analyzing gait with spatiotemporal surfaces. In *Motion of Non-Rigid and Articulated Objects, Proceedings of the IEEE Workshop on*, pp. 64–69.

Oh, S., A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J. Lee, S. Mukherjee, J. Aggarwal, H. Lee, L. Davis, E. Swears, X. Wang, Q. Ji, K. Reddy, M. Shah, C. Vondrick, H. Pirsiavash, D. Ramanan, J. Yuen, A. Torralba, B. Song, A. Fong, A. Roy-Chowdhury, and M. Desai (2011). A large-scale benchmark dataset for event recognition in surveillance video. In *Computer Vision and Pattern Recognition, Proceedings of the IEEE Conference*.

Ojala, T., M. Pietikäinen, and T. Mäenpää (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 24*(7), 971–987.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing 28*(6), 976 – 990.

Reddy, K. and M. Shah (2013). Recognizing 50 human action categories of web videos. *Machine Vision and Applications 24*(5), 971–981.

Rodriguez, M., J. Ahmed, and M. Shah (2008). Action mach: a spatio-temporal maximum average correlation height filter for action recognition. In *Computer Vision and Pattern Recognition, Proceedings of IEEE International Conference on*.

Rudoy, D. and L. Zelnik-Manor (2012). Viewpoint Selection for Human Actions. *International Journal of Computer Vision 97*(3), 243–254.

Rvachev, V. L. (1982). *Theory of R-functions and Some Applications*. Naukova Dumka. Russian.

Sarkar, S., P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer (2005). The humanid gait challenge problem: data sets, performance, and analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 27*(2), 162–177.

Scharr, H., S. Körkel, and B. Jähne (1997). Numerische Isotropieoptimierung von FIR-Filtern mittels Querglättung. In *Proceedings of DAGM*, pp. 367–374.

Scholkopf, B. and A. Smola (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.

Schuldt, C., I. Laptev, and B. Caputo (2004). Recognizing human actions: a local SVM approach. In *Pattern Recognition, Proceedings of the International Conference on*, Volume 3, pp. 32 – 36.

Seely, R. (2010). *On a three-dimensional gait recognition system*. Ph. D. thesis, School of Electronics and Computer Science, University of Southampton.

Seely, R., M. Goffredo, J. Carter, and M. Nixon (2009). View invariant gait recognition. In *Handbook of Remote Biometrics: for Surveillance and Security*, pp. 61–82. Springer.

Seely, R., S. Samangooei, L. Middleton, J. Carter, and M. Nixon (2008). The university of southampton multi-biometric tunnel and introducing a novel 3D gait dataset. In *Biometrics: Theory, Applications and Systems*.

Sethi, M., R. A., and K. S. Gurumoorthy (2012). The Schrödinger distance transform (SDT) for point-sets and curves. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 198–205.

Shah, J. (1991). Segmentation by nonlinear diffusion. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 202–207.

Shapiro, V. (2007). Semi-analytic geometry with R-functions. *Acta Numerica 16*, 239–303.

Singh, S., S. Velastin, and H. Ragheb (2010). MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods. In *Advanced Video and Signal Based Surveillance, IEEE International Conference on*, pp. 48–55.

Sivapalan, S., D. Chen, S. Denman, S. Sridharan, and C. Fookes (2011). Gait Energy Volumes and Frontal Gait Recognition using Depth Images. In *International Joint Conference on Biometrics*, pp. 1–6.

Spalding, D. B. (1994). Calculation of turbulent heat transfer in cluttered spaces. In *Heat Transfer Conference, Proceedings of the International*, Brighton, UK.

Sun, B., J. Yan, and Y. Liu (2010). Human gait recognition by integrating motion feature and shape feature. In *Multimedia Technology, International Conference on*, pp. 1 – 4.

Tan, D., K. Huang, S. Yu, and T. Tan (2006). Efficient night gait recognition based on template matching. In *Pattern Recognition, International Conference on*, Volume 3, pp. 1000 – 1003.

Tari, Z. S. G., J. Shah, and H. Pien (1997). Extraction of shape skeletons from grayscale images. *Computer Vision and Image Understanding 66*(2), 133–146.

Tran, D. and A. Sorokin (2008a). Human activity recognition with metric learning. Volume 5302 of *Lecture Notes in Computer Science*, pp. 548–561. Springer Berlin Heidelberg.

Tran, D. and A. Sorokin (2008b). Human activity recognition with metric learning. In *European Conference on Computer Vision Part I, Proceedings of the*, pp. 548 – 561.

Troje, N. (2002). Biomotion lab. http://www.biomotionlab.ca. Accessed April 2014.

Tucker, P. G. (1998). Assessment of geometric multilevel convergence and a wall distance method for flows with multiple internal boundaries. *Applied Mathematical Modelling 22*, 293–311.

UCF. University of Central Florida aerial camera, rooftop camera and ground camera dataset. http://crcv.ucf.edu/data/UCF-ARG.

UCF. University of Central Floridaaerial action dataset. http://crcv.ucf.edu/data/UCF_Aerial_Action.

van der Maaten, L. (2007). Matlab Toolbox for Dimensionality Reduction.

Varadhan, S. R. S. (1967). On the behavior of the fundamental solution of the heat equation with variable coefficients. *Communications of Pure and Applied Mathematics 20*, 431–455.

Veeraraghavan, A., A. Chowdhury, and R. Chellappa (2004). Role of shape and kinematics in human movement analysis. In *Computer Vision and Pattern Recognition, Proceedings of the IEEE Computer Society Conference on*, Volume 1, pp. I–730–I–737.

Veres, G., L. Gordon, J. Carter, and M. Nixon (2004). What image information is important in silhouette-based gait recognition? In *Computer Vision and Pattern Recognition, Proceedings of the IEEE Computer Society Conference on*, Volume 2, pp. II–776–II–782.

ViSOR (2011). The imagelab laboratory of the university of modena and reggio emilia, visor (video surveillance online repository). http://www.openvisor.org/index.asp.

Wang, C., J. Zhang, J. Pu, X. Yuan, and L. Wang (2010). Chrono-gait image: A novel temporal template for gait recognition. Volume 6311, pp. 257–270. Springer Berlin Heidelberg.

Wang, C., J. Zhang, L. Wang, J. Pu, and X. Yuan (2012). Human Identification Using Temporal Information Preserving Gait Template. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 34*(11), 2164–2176.

Wang, H., M. Ullah, A. Klaser, I. Laptev, and C. Schmid (2009). Evaluation of local spatio-temporal features for action recognition. In *British Machine Vision Conference*.

Wang, J., M. She, S. Nahavandi, and A. Kouzani (2010). A review of vision-based gait recognition methods for human identification. In *Digital Image Computing: Techniques and Applications, International Conference on*, pp. 320–327.

Wang, L., W. Hu, and T. Tan (2003). Recent developments in human motion analysis. *Pattern Recognition 36*(3), 585 – 601.

Wang, L., H. Ning, T. Tan, and W. Hu (2004). Fusion of static and dynamic body biometrics for gait recognition. *Circuits and Systems for Video Technology, IEEE Transactions on 14*(2), 149–158.

Wang, L., T. Tan, H. Ning, and W. Hu (2003). Silhouette analysis-based gait recognition for human identification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 25*(12), 1505–1518.

Wang, Y. and G. Mori (2009). Max-margin hidden conditional random fields for human action recognition. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 872 – 879.

Weickert, J. and H. Scharr (2002). A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. *Journal of Visual Communication and Image Representation 13*, 103–18.

Weinland, D., M. Özuysal, and P. Fua (2010). Making action recognition robust to occlusions and viewpoint changes. In *European Conference on Computer Vision*.

Weinland, D., R. Ronfard, and E. Boyer (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding 104*(23), 249 – 257.

Weinland, D., R. Ronfard, and E. Boyer (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding 115*(2), 224 – 241.

Whytock, T., A. Belyaev, and N. Robertson (2012). GEI + HOG for action recognition. In *British Machine Vision Conference Student Workshop*.

Whytock, T., A. Belyaev, and N. Robertson (2013a). Improving Robustness and Precision in GEI + HOG Action Recognition. In *Advances in Visual Computing*, Volume 8033 of *Lecture Notes in Computer Science, Part I*, pp. 119–128. Springer. (Presented at the International Symposium on Visual Computing).

Whytock, T., A. Belyaev, and N. Robertson (2013b). Robust gait recognition via covariate factor mitigation. In *Imaging for Crime Detection and Prevention, International Conference on*.

Whytock, T., A. Belyaev, and N. Robertson (2013c). Towards Robust Gait Recognition. In *Advances in Visual Computing*, Volume 8034 of *Lecture Notes in Computer Science, Part II*, pp. 523–531. Springer. (Presented at the International Symposium on Visual Computing).

Whytock, T., A. Belyaev, and N. Robertson (2014). Dynamic distance-based shape features for gait recognition. *Journal of Mathematical Imaging and Vision 50*(3), 314–326.

Whytock, T., A. Belyaev, and N. Robertson (2015). On covariate factor detection and removal for robust gait recognition. *Machine Vision and Applications*, 1–14.

Willems, G., T. Tuytelaars, and L. Gool (2008). An efficient dense and scale-invariant spatio-temporal interest point detector. In *European Conference on Computer Vision Part II, Proceedings of the*, pp. 650 – 663.

Wren, C., A. Azarbayejani, T. Darrell, and A. Pentland (1997). Pfinder: real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 19*(7), 780–785.

Xia, H., P. G. Tucker, and G. Coughlin (2012). Novel applications of BEM based Poisson level set approach. *Engineering Analysis with Boundary Elements 36*, 907–912.

Yam, C., M. Nixon, and J. Carter (2004). Automated person recognition by walking and running via model-based approaches. *Pattern Recognition 37*(5).

Yam, C.-Y. and M. Nixon (2009a). *Encyclopedia of Biometrics*, Chapter Model-based Gait Recognition, pp. 1082–1088. Springer.

Yam, C.-Y. and M. Nixon (2009b). Model-based gait recognition. In *Enclycopedia of Biometrics*, pp. 633–639. Springer.

Yeffet, L. and L. Wolf (2009). Local trinary patterns for human action recognition. In *Computer Vision, IEEE International Conference on*, pp. 492 – 497.

Yogarajah, P., J. Condell, and G. Prasad (2011). P_{RW}GEI: Poisson Random Walk based Gait Recognition. In *Image and Signal Processing and Analysis, International Symposium on*, pp. 662 – 667.

Yoo, J. and M. Nixon (2011). Automated Markerless Analysis of Human Gait Motion for Recognition and Classification. *ETRI Journal 33*(3), 259–266.

Yu, S., D. Tan, and T. Tan (2006a). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Pattern Recognition, International Conference on*, Volume 4, pp. 441–444.

Yu, S., D. Tan, and T. Tan (2006b). A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In *Pattern Recognition, International Conference on*, Volume 4, pp. 441–444.

Yu, S., T. Tan, K. Huang, K. Jia, and X. Wu (2009). A study on gait-based gender classification. *Image Processing, IEEE Transactions on 18*, 1905 – 1910.

Yuan, J., Z. Liu, and Y. Wu (2009). Discriminative subvolume search for efficient action detection. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pp. 2442–2449.

Zhang, E., Y. Zhao, and W. Xiong (2010). Active energy image plus 2DLPP for gait recognition. *Signal Processing 90*(7), 2295 – 2302.

Zheng, S., J. Zhang, K. Huang, R. He, and T. Tan (2011). Robust View Transformation Model For Gait Recognition. In *Image Processing, IEEE International Conference on*, pp. 2073–2076.